



LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

Published:

— *Without international search report and to be republished upon receipt of that report.*

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
11 January 2001 (11.01.2001)

PCT

(10) International Publication Number
WO 01/02553 A3

(51) International Patent Classification⁷: C12N 15/63,
15/85, C07K 14/46, C12N 9/10, A61K 48/00

Street, Kensington, MD 20985 (US). **FELSENFELD, Gary** [US/US]; 3109 Leland Street. Chevy Chase, MD 20815 (US).

(21) International Application Number: PCT/US00/10509

(74) Agents: **FEILER, William, S.** et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).

(22) International Filing Date: 19 April 2000 (19.04.2000)

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(26) Publication Language: English

(30) Priority Data:
60/141,728 30 June 1999 (30.06.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/141,728 (CIP)
Filed on 30 June 1999 (30.06.1999)

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— With international search report.

(71) Applicant (*for all designated States except US*): **THE GOVERNMENT OF THE UNITED STATES OF AMERICA** as represented by **THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES** [US/US]; Office of Technology Transfer, National Institutes of Health, Suite 325, 6011 Executive Boulevard, Rockville, MD 20852 (US).

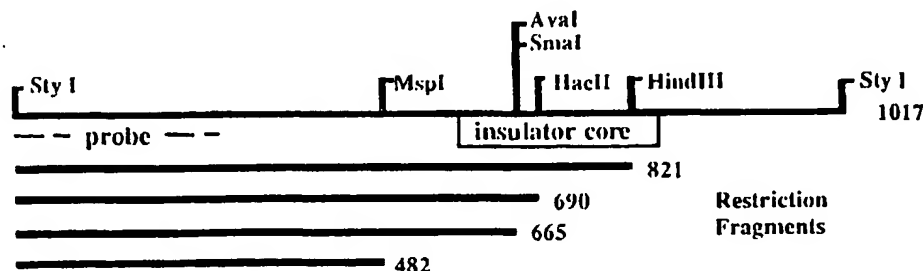
(88) Date of publication of the international search report:
3 May 2001

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **BELL, Adam, C.** [US/US]; 18670 Winding Creek Place, Germantown, MD 20874 (US). **WEST, Adam, G.** [US/US]; 4511 Everett

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DNA BINDING PROTEIN AND SEQUENCE AS INSULATORS HAVING SPECIFIC ENHANCER BLOCKING ACTIVITY FOR REGULATION OF GENE EXPRESSION



(57) Abstract: An insulator element shown to be both necessary and sufficient for the enhancer blocking activity in human cells is described. The insulator element of the invention is a 42 base pair DNA molecule having specific enhancer blocking activity for regulation of gene expression. The insulator element has also been shown to be the core binding site for CTCF, a DNA binding protein that is highly conserved in vertebrates. In addition, an insulator element containing CTCF binding sites has been identified in the region of the Igf 2 locus which is methylated exclusively on the paternal allele. Methylation of the insulator sequence abolishes the ability of CTCF to bind to the insulator and results in loss of enhancer-blocking activity.

WO 01/02553 A3

Title Of The Invention

5 DNA BINDING PROTEIN AND SEQUENCE
AS INSULATORS HAVING SPECIFIC ENHANCER BLOCKING
ACTIVITY FOR REGULATION OF GENE EXPRESSION

Field Of The Invention

10 The present invention relates to the identification of an insulator
nucleic acid sequence which has the ability to block the action of enhancers and
prevent gene activation, and a DNA binding protein which binds to the insulator
sequence. The invention also relates to methods for insulating the expression of a
given gene by employing the insulator sequence and/or the DNA binding protein of
the invention. The invention further relates to the identification of an insulator
15 element which has the ability to block the expression of the insulin growth factor 2
(Igf2) gene. The enhancer-blocking activity of this insulator element is dependent
upon CTCF binding to the insulator. Methylation of the insulator element abolishes
the ability of the CTCF to bind to the insulator and would therefore result in loss of
CTCF-dependent enhancer-blocking activity. The invention also relates to methods
20 of modulating the enhancer-blocking activity of the insulator element.

Background Of The Invention

Enhancer-mediated activation is a fundamental mechanism of gene
regulation in eukaryotic organisms. Enhancers can act over large distances to
25 activate transcription, independent of their orientation and position relative to the
promoter. In many cases, if given access, enhancers can act promiscuously to
activate transcription of heterologous promoters. In fact, some types of cancers are
thought to arise as a result of translocations which artificially juxtapose an oncogene
with a heterologous enhancer.

30 Genome sequencing has revealed many cases where differentially
regulated genes neighbor each other at distances over which enhancers could act,
yet the genes are independently regulated. Thus, mechanisms are likely to exist that
are able to prevent the action of an enhancer on a neighboring locus. This
restriction must be achieved, at least in some cases, without impeding the action of
35 the enhancer within its native locus. A DNA element able to function in this way

- 2 -

would, in effect, constitute a boundary to the action of an enhancer, thereby preventing it from acting across the boundary, while otherwise leaving the enhancer unimpeded. This property is one of the defining characteristics of an insulator, a type of regulatory element that has only recently been recognized. (Kellum and
5 Elgin, 1998; Udvardy, 1999; Bell and Felsenfeld, 1999, U.S. Patent No. 5,610,053 to J. Chung et al.).

The first DNA sequences to be described as having the properties of an insulator were the *scs* and *scs'* elements of *Drosophila*, which were initially identified as marking the chromatin boundaries of a heat shock locus. When *scs*
10 elements (i.e., *scs* DNA sequences) were placed on either side of a gene for eye color and introduced as transgenes into *Drosophila* embryos, the resulting offspring flies all had similar eye color, independent of the site of integration of the transgene. This result indicated that *scs* had protected the reporter gene from both negative and positive endogenous influences, or 'position effects' (Kellum and Schedl, 1991 and
15 1992). Another *Drosophila* insulator element, *gypsy*, was first identified because of its ability to block the action of an enhancer on a promoter when the element lay between them, but not otherwise (Holdridge and Dorsett 1991; Geyer and Corces, 1992; Dorsett, 1993). Studies of these elements have led to a working definition of an insulator as an element that is capable of protecting against position effects
20 and/or blocking enhancer action in a directional manner. For both *scs'* and *gypsy*, proteins have been identified that bind specifically to the DNA elements and are, at least in part, responsible for mediating insulator activity (Geyer and Corces, 1992; Zhao et al., 1998).

Insulator elements have also been identified in vertebrates (Chung et
25 al., 1993 and 1997; Zhong and Krangel, 1997; Robinett et al., 1997). U.S. Patent No. 5,610,053 to Chung et al. has described a 1.2 kb DNA insulator element, which was derived from the 5' end of the chicken β -globin locus and exhibited strong enhancer-blocking activity. (Chung et al., 1993 and 1997). This region contains a constitutive DNase I hypersensitive site that is present in all tissues. The 1.2 kb
30 insulator element coincides almost exactly with the point of transition between an active chromatin conformation, marked both by histone hyperacetylation and a

- 3 -

heightened general sensitivity to DNase, and an inactive domain extending farther 5' that is insensitive to nuclease and less highly acetylated. (Hebbes et al., 1994).

Within the 1.2 kb element is a 250 base pair (bp) 'core' fragment or region that possesses a large part of the enhancer blocking activity (U.S. Patent No. 5,610,053 to J. Chung et al.; Chung et al., 1997). However, the identification and characterization of additional and specific sub-sequences having insulator function within the 1.2 kb insulator element and 250 bp core remain to be elucidated. In addition, there remains to be discovered and identified one or more DNA binding sites within the core region that is/are necessary and sufficient for enhancer-blocking activity and that are recognition sites for regulatory protein binding to DNA.

Directional enhancer-blocking activity of proteins that bind to specific insulator nucleic acid sequences provides to the art important methods to control gene function at numerous complex gene loci in many organisms. The identification and characterization of an insulator sequence and protein that binds thereto can establish the foundation for maintenance of boundaries between different groups of genes that have distinct regulatory patterns. The use of such isolated sequences and their purified binding proteins provide significant tools for the regulation of gene expression and function in mammals and plants.

Description Of The Drawings

Figs. 1A-1G show the fine mapping of the insulator core. In Figs. 1A and 1B, the position of HS4 was measured by comparing the migration of the DNase I digestion fragment generated by limited digestion of chicken erythrocyte chromatin to the migration of DNAs of known length and identical composition. The logic of this mapping is outlined in Fig. 1A. Fig. 1B shows the autoradiographic results of enhancer blocking activity of fragments of the core. The position of the hypersensitive site relative to previously defined DNase I footprints (Chung et al., 1997) is indicated in Fig. 1C. Figs. 1D-1F show the results of enhancer blocking assays in which the elements indicated were placed between enhancer and promoter as depicted in Fig. 1G, and the relative number of neomycin-resistant colonies was counted. A schematic of each inserted element is shown

- 4 -

("Test Fragments"), as well as the relative numbers of neomycin-resistant colonies observed ("Relative Neo^R Colonies"), and the numerical value of the insulation effect ("Fold Insulation") relative to the non-insulated controls (pNI and λ DNA). In particular, Figs. 1D and E show the effect on enhancer blocking of deletion of footprinted regions from the core. Fig. 1F shows increased enhancer blocking that is observed when insulating elements were multimerized. The data presented in Figs. 1D-F represent an average of at least 4 independent assays. Fig. 1G presents a schematic diagram of the construction used to test various DNA fragments for enhancer-blocking activity.

Figs. 2A-2E present the identification of a minimal enhancer blocking site. In Fig. 2A, a 90 bp fragment spanning FII and FIII was subjected to further deletion; the effects on enhancer blocking are shown. Fig. 2B presents the results of an examination of the effect of the relative positions of enhancer and promoter on the enhancer-blocking effect of FII in the colony assay. Fig. 2C presents the effect of mutation of Sp1 sites within FII/III on enhancer blocking. Fig. 2D shows the results of an evaluation of the contribution to the enhancer blocking activity of FII of motifs homologous to $\alpha 2$ and Su(Hw). Fig. 2E shows a truncated sequence of FII (SEQ ID NO:2) and its homologies to known transcription factor binding sites, namely, Su(Hw): SEQ ID NO:3; Sp1; SEQ ID NO:4; and $\alpha 2$: SEQ ID NO:5. The enhancer-blocking data presented in Figs. 2A-D represent the average of 2-5 independent assays for each construction.

Figs. 3A-3D show the sequence specificity of enhancer blocking and nuclear factor binding by FII. In Fig. 3A, the indicated sequences (SEQ ID NOS:1, 6-10) were inserted into the AscI site of pNI; their ability to block an enhancer was measured in the colony assay. Specifically, as shown in Fig. 3A, SEQ ID NO:1 corresponds to the FII fragment; SEQ ID NO:6 corresponds to x5'; SEQ ID NO:7 corresponds to xM; SEQ ID NO:8 corresponds to x3'; SEQ ID NO:9 corresponds to ΔF ; and SEQ ID NO:10 corresponds to rev sequences. Figs. 3B and 3C show gel mobility shift assays with a labeled 60 bp FII probe and nuclear extracts from human K562 (Fig. 3B) and chicken red blood cells (RBC), (Fig. 3C). Cold competitors as shown (sequences in Fig. 3A) were added at a 100 fold molar excess

– 5 –

in these experiments. Fig. 3D presents a comparison of the capacity of the indicated FII mutants (SEQ ID NOS:1, 6-17) to (i) act as insulators in the colony assay, (ii) bind to a candidate insulator protein in gel shift and (iii) bind to CTCF in southwestern binding assays (see also Figs. 4A-C). Data were normalized with FII activity considered as 100% in each assay.

Figs. 4A-4D relate to purification of an FII binding factor. Fig. 4A shows sequence specific FII binding observed at ~140 kDa apparent molecular weight in protein fractions obtained during different stages of purification. Fig. 4B shows a schematic outline of the protocol used to purify the FII binding factor.

Figs. 4C and 4D provide a representative example of a Coomassie stained gel of the purified fractions eluted from the hydroxyapatite column with the internal peptide sequences obtained from the indicated band indicated (labeled "Coomassie", Fig. 4C), and the result of a southwestern assay of FII binding to this fraction (labeled "Southwestern", Fig. 4D).

Fig. 5 shows FII binding and enhancer blocking by CTCF. Purified FII binding factor (lane 1) and *in vitro* translated CTCF (lanes 2-11) have identical specificity for FII (compare to Fig. 3A) and identical complex migration in a gel shift assay when bound to either FII (lanes 2-8) or previously-characterized CTCF sites from the chicken *c-myc* promoter (lane 9), the chicken *lysozyme* promoter (lane 10) or the human *amyloid beta-protein* promoter (lane 11). The table in the right of the figure summarizes the capacity of CTCF sites to act as enhancer-blockers in the colony assay. The data presented represent the average of two independent measurements.

Figs. 6A-6C show sequence homologies among CTCF sites and vertebrate insulators. Fig. 6A, (SEQ ID NOS:1, 8, 18-20), shows that the alignment of FII with other known CTCF sites reveals a conserved 3' region which corresponds to the sequence altered in the x3' mutant (see Figs. 3A-3D). Fig. 6B, (SEQ ID NOS:1, 21-27), shows the alignment of the 100 bp repeats of the *Xenopus* RO element and FII. Fig. 6C, (SEQ ID NOS:1 and 28), shows the alignment of FII with a homologous site (BEAD-A) in the BEAD-1 element from the human T cell receptor α/δ locus.

- 6 -

Figs. 7A and 7B show the conservation of sequence-specific enhancer blocking activity among vertebrate insulators. Fig. 7A presents gel mobility shift assays with FII and BEAD A as probes to reveal sequence specific binding to partially purified CTCF. An antibody raised against a C-terminal peptide of CTCF specifically supershifts both complexes. Fig. 7B shows enhancer blocking activities of vertebrate insulators. These data are the average of at least two independent experiments, with the exception that the data for the RO element are from a single determination.

Figure 8A-B show that differentially methylated region (ICR) upstream of *H19* has the enhancer-blocking properties of an insulator. Fig. 8A shows a schematic of the neighboring mouse *Igf2* and *H19* genes. On the maternally inherited chromosome, the ICR is unmethylated (white rectangles) and contains two nuclease-hypersensitive regions (hatched boxes, HS1 and HS2); on the paternally inherited chromosome, the ICR is methylated (black rectangles) and contains no hypersensitive sites. Deletion of a 1.6 kb fragment of the ICR (termed the DMD fragment) eliminates HS2 and most of HS1. Fig. 8B shows enhancer-blocking activity of various constructs. Constructs in which various fragments of the ICR were inserted at defined positions relative to the enhancer and promoter were prepared. For each construct, colony number was normalized to an uninsulated control, NI. Data are the average of three independent measurements.

Figure 9A-D shows conserved CTCF sites within the *H19* ICR. Fig. 9A shows sequences of the CTCF sites clustered upstream of the mouse, rat, and human *H19* genes. Shading indicates identity among the sites; gray shading indicates species-specific identities, while black shading indicates cross-species sequence conservation among these sites. Fig. 9B shows enhancer-blocking activity of a fragment spanning only m3 from the mouse ICR. Data are the average of three independent experiments. Fig. 9C shows gel mobility-shift analysis of β -globin FII (60-mer) and 83-91-mer duplexes spanning mouse and human ICR sites binding to K562 nuclear extract (E), partially purified (chicken) CTCF (P), and *in vitro* translated human CTCF (I). An asterisk indicates the position of the CTCF:DNA complex. Labeled DNA probes are indicated at the panel bottom. Fig. 9D shows analysis of CTCF binding to representative mouse and human ICR sites. DNAs

- 7 -

were incubated with K562 nuclear extract in the presence of a 50-fold excess of unlabeled competitors as indicated or with an anti-CTCF antibody. SS indicates the position of the supershifted CTCF complex.

Figures 10A-B show that CTCF is responsible for the methylation-sensitive enhancer-blocking activities of the mouse and human ICRs. Fig. 10A shows enhancer blocking activities of fragments of the mouse ICR after sequential deletion of the sequences spanning individual CTCF sites. Results are the average of two to three independent measurements. Fig. 10B shows on the left panel the effect of CpG methylation on binding of partially purified chicken CTCF to various sites in the absence of competitor DNA (-), in the presence of 50-fold excess of unlabeled duplex DNA (S, self) or a 50-fold excess of unlabeled duplex of identical sequence with ^{5me}C incorporated at every CpG (M, uniformly methylated). Labeled DNA probes are indicated at the bottom of the panel. In the right panel: the effect of ^{5me}C substitution at a single site (M1, singly methylated at the first CpG in the black-shaded region of Fig. 9a).

Figure 11 shows a model for methylation-dependent modulation of insulator action in the epigenetic regulation of *Igf2*. On the maternally inherited chromosome, the ICR is unmethylated. This allows binding of CTCF to its sites (m1-4), two in each nuclease-hypersensitive region (shaded boxes), and the resulting insulator blocks activation of the maternal copy of *Igf2* by the *H19* enhancer. On the paternally inherited chromosome, the ICR is methylated. This prevents CTCF binding, thereby inactivating the insulator and allowing the *H19* enhancer to activate *Igf2*.

Summary Of The Invention

The present invention provides a newly-identified insulator nucleic acid sequence that acts as a barrier to the influences of neighboring *cis*-acting elements, thereby preventing gene activation, for example, when juxtaposed between an enhancer sequence and a promoter sequence. According to the present invention, the new insulator nucleic acid sequence is 42 base pairs (bp) in length and comprises a new, specific and previously unidentified fragment of the chicken beta (β)-globin insulator element. This insulator sequence was shown to be both

– 8 –

necessary and sufficient for the enhancer blocking activity in human cells, as described herein.

5 It is an object of the present invention to provide a method for using the newly-characterized and isolated insulator element to insulate or buffer the expression of a reporter gene from adverse effects of neighboring or surrounding chromatin. The incorporation of the defined insulator sequence into vectors and constructs allows gene transfer and expression in cells and tissues with virtually no concern for suppression or inhibition of expression due to the chromosomal milieu after integration.

10 It is another object of the present invention to provide genetic expression constructs or vectors which are designed to contain one or more operational DNA sequence insulator elements comprising SEQ ID NO:1 which can insulate or buffer the activity of a particular gene from the effects of the activity of *cis*-acting regulatory elements, such as enhancer or silencer regions of the DNA.

15 The constructs may contain one or more insulator elements and one or more reporter genes in the form of transcription units or mini-loci, including at a minimum, an enhancer, a promoter, and a reporter gene. The insulator element-containing constructs allow for the transfection of cells of a particular lineage or of a particular tissue type, depending upon the gene to be transfected and upon other features of the

20 construct which may be cell- or tissue-specific, such as specific promoter or enhancer elements, or upon particular regulatory molecules, proteins, or factors which are produced by a particular cell or tissue type and which influence the expression of a given transfected gene.

In accordance with the invention, the insulator element(s), reporter

25 gene(s), and transcription unit may be provided in the form of a cassette designed to be conveniently ligated into a suitable plasmid or vector, which plasmid or vector is then used to transfect cells or tissues, and the like, for both *in vitro* and *in vivo* use.

It is a further object of the present invention to provide a mechanism and a tool to restrict the action of *cis*-acting regulatory elements on genes whose

30 activities or encoded products are needed or desired to be expressed in certain cells and tissues. The genes to be insulated and expressed may be introduced into cells by employing the constructs or vectors achieved by the present invention in which

– 9 –

one or more insulator elements in a chromatin domain are strategically positioned so as to buffer the transfected genes from the influence of the action of other DNA sequences from different chromatin domains located in *cis*.

It is also an object of the present invention to provide a specific
5 binding site for a purified protein, CTCF, which is an eleven zinc finger DNA binding protein, highly conserved in vertebrates. The sequence specificity of CTCF accounts precisely for the sequence requirements of directional enhancer-blocking *in vivo*.

It is a further object of the present invention to provide a method for
10 insulating a given gene by employing the insulator sequence and/or the CTCF binding protein to achieve directional enhancer blocking of a gene.

Also provided is a kit or kits containing the vector constructs of the invention and used to insulate the expression of a heterologous gene or genes integrated into host DNA.

15 The invention further provides a method and constructs to insulate the expression of a gene or genes in transgenic animals such that the transfected genes will be able to be protected and stably expressed in the tissues of the transgenic animal or its offspring, for example, even if the DNA of the construct integrates into areas of silent or active chromatin in the genomic DNA of the host
20 animal.

Yet another object of the present invention to provide a method for insulating the expression and function of a given gene by employing the DNA binding protein CTCF to bind to the insulator sequence as described herein.

The invention further relates to the identification of an insulator
25 element which has the ability to block the expression of the insulin growth factor 2 (Igf2) gene. This insulator element contains CTCF binding sites and its enhancer-blocking activity is dependent upon CTCF binding to these sites. Methylation of the insulator element abolishes the ability of the CTCF to bind to the insulator and would therefore result in loss of the CTCF-dependent enhancer-blocking activity.

30 The invention also relates to methods of modulating the enhancer-blocking activity of an insulator by targeted methylation or demethylation of the insulator.

– 10 –

Further objects and advantages of the present invention will be appreciated in light of the description herein.

Detailed Description Of The Invention

5 The present invention provides an isolated, 42 base pair (bp) fragment (DNA sequence motif) of the chicken β -globin insulator which has been newly found to be both necessary and sufficient for enhancer blocking activity in human cells. This DNA fragment, called FII herein, has been found to serve as an insulator molecule, i.e., a DNA sequence which can act as a barrier to the influences
10 of neighboring *cis*-acting elements, for example, to prevent gene activation when located between an enhancer and a promoter of a given gene.

According to the present invention, this small DNA sequence motif, FII, comprises the minimal binding site for a cellular DNA binding protein and has the following DNA sequence, from 5' to 3' :

15 5' - CCCAGGGATGTAATTACGTCCCTCCCCGCTA
GGGGGCAGCA - 3' (SEQ ID NO:1), (Figs. 3A, 3D and 6A-6C).

Newly identified and isolated in accordance with the present invention, the sequence motif of SEQ ID NO:1 accounts for most of the ability of the insulator element to block enhancer activity. Indeed, the function of this smaller
20 sequence has been shown to be pivotal to the ability of the insulator element to block the action of enhancers.

The 42 bp fragment containing the sequence motif of SEQ ID NO:1 was able to suppress enhancer activity in a directional manner about as well as the full 1.2 kb element as described in U.S. Patent No. 5,610,053 to J. Chung et al., the
25 contents of which are hereby incorporated by reference herein. The fragment also contains binding sites for Sp1 and the yeast $\alpha 2$ repressor; however, mutation of these sites had no effect on its blocking activity. By contrast, mutations of the 3' end of FII site did abolish enhancer-blocking.

According to the present invention, a DNA fragment encompassing
30 SEQ ID NO:1 has also been newly discovered to be the core binding site for CTCF, a DNA binding protein that is highly conserved in vertebrates. A further significance of the CTCF site is that it has been found within the BEAD 1 element

- 11 -

of the T-cell α/δ locus (Zhong and Krangel, 1997) and accounts for the activity of the BEAD 1 element. The DNA sequence motif characterized by SEQ ID NO:1 specifies the minimal functional binding site for CTCF, as demonstrated in *in vitro* studies. For optimal binding by the CTCF protein, the DNA sequence motif

5 characterized by SEQ ID NO:1 preferably comprises additional base pairs at the the 5' and 3' ends, thereby yielding a DNA binding element having approximately 50 to 70, more preferably 50-60 total base pairs, which includes SEQ ID NO:1. For example, the DNA binding element for CTCF may comprise SEQ ID NO:1 and from about 5-20 additional bases, added to the 5' and 3' ends of the sequence,

10 without adversely affecting its insulator function.

When the 42 bp FII DNA fragment was used as a probe in gel retardation experiments, two major shifted bands were observed. One of these bands was attributable to interaction with Sp1. The other band appeared to be associated with insulating activity, based on its properties determined from

15 competition experiments. More specifically, the second of the two bands was competed by any DNA that was also active in the enhancer blocking assay, but not by any of the inactive mutated sequences that were tried. Similar gel shift patterns were obtained with extracts from nuclei from chicken erythrocytes and from the human erythroleukemia cell line, K562, which were used to carry out the enhancer

20 blocking assays. These observations and results were used to purify the protein responsible for the specific shifted band. The purified protein was determined to be the protein CTCF. The product of an *in vitro* transcription/translation reaction with cloned CTCF cDNA yielded results that were identical to those obtained using the above-mentioned cell extracts.

25 CTCF is an 82 kDa protein with 11 zinc fingers (Filippova et al. 1996), and is characterized by an unusually extensive DNase I footprint (51 bp) when bound to its site on DNA, consistent with an involvement of several fingers in typical binding sites. It migrates aberrantly on acrylamide/SDS gels, which accounts for the discrepancy in apparent molecular weight (Klenova et al., 1997).

30 Studies of CTCF in other systems suggest that it can play a variety of regulatory roles. For example, it binds to the promoter of the amyloid β -protein precursor and causes transcriptional activation (Vostrov and Quitschke, 1997), but when it

– 12 –

interacts with sites in the *c-myc* oncogene, it causes repression (Filippova 1996). CTCF has also been found to be capable of acting in synergy with certain thyroid hormone receptor binding sites both in repression and in T3 induction (Baniahmad et al., 1990). Not all of the 11 zinc fingers of the protein are involved in binding to the sites that have been examined so far. Furthermore, different sites employ partially different subsets of fingers to contact the DNA (Filippova et al. 1996). From this, it can be surmised that the characteristics of the binding site would have a large influence on the conformation of the protein, the nature of its interactions with cofactors, and its ultimate biological effect(s).

10 According to the results presented herein, the CTCF binding site is necessary and sufficient for enhancer blocking activity, as demonstrated in the exemplified assays. The presence of similar binding sites at each of the vertebrate loci known to have enhancer-blocking activity is strong evidence for the role of CTCF sites in insulator function *in vivo*.

15 Also according to the present invention, the CTCF binding site has been determined to be a sub-fragment (42 bp) of the larger 1.2 kb insulator element containing the β -globin 5' HS4. As mentioned above, it is noteworthy that the β -protein 5' insulator element shares with the *Drosophila* insulators the additional ability to protect against position effects. For example, when two copies of the entire 1.2 kb fragment containing the β -globin 5' HS4 are placed on either side of a stably-integrated reporter gene, the reporter is protected both against variation in expression from one line to another and also against extinction of expression over a period of at least 40-80 days in culture (Pikaart et al., 1998). It is likely that this activity depends upon sequences other than or in addition to FII within the larger insulator element. The complete activity of the β -globin 5' insulator element is thus likely to involve multiple components.

25 It has been newly determined as described herein that enhancer blocking activity of the 5' β -globin insulator and the isolated portions thereof, is dependent upon CTCF. In addition, similar DNA binding site sequences are present in two other vertebrate insulators. The first of these is the BEAD-1 element found in the human T-cell receptor (TCR) α/δ locus (Zhong and Krangel, 1997). BEAD-

30

- 13 -

1, which has strong directional enhancer blocking properties, is located between TCR δ gene segments and TCR α joining gene segments. It has been proposed that BEAD-1 prevents a δ -specific enhancer from acting on the α genes early in T-cell development. The present findings have shown that BEAD-1 contains a CTCF
5 binding site (i.e., the 42 bp DNA sequence motif sub-fragment of the 1.2 kb element) and that this site is responsible for a large portion of the observed enhancer blocking activity.

Taken together, the results described herein suggest a conserved and perhaps widely used function of insulators in which CTCF is involved in the
10 maintenance of distinct regulatory regions. Indeed, additional analyses in the inventors' laboratory have shown that the 3' end of the chicken β -globin locus is marked by a hypersensitive site with similar properties to 5' HS4. This 3' end hypersensitive site also contains a CTCF binding site. This element is located between the globin genes and a nearby, yet distinctly unrelated, gene encoding an
15 odorant receptor (Burger et al., 1999), further substantiating the nature and likely function of these boundary elements *in vivo*.

Insulators typically are capable of both blocking enhancer activity and protecting against position effects. These two functions might have only partially overlapping mechanisms. Protection against position effects implies that
20 activation by external endogenous enhancers is blocked, consistent with the activity described herein. However, position effects also arise from silencing induced by neighboring heterochromatin. While the insulator described herein is able to protect against external position effects, it may also be that additional components of the insulator element, or additional cofactors, are involved in protecting against such
25 effects.

Without wishing to be bound by theory, it may be likely that in some situations, where enhancer blocking activity is all that is required, a CTCF binding site alone is sufficient, while in the case of a permanent chromatin domain boundary, such as that found at the 5' end of the chicken β -globin locus, additional
30 components are involved. Indeed, even in those cases where only CTCF sites are present, the activity of CTCF may require the participation of other proteins, just as

– 14 –

the directional enhancer-blocking activity of the suppressor of Hairy-wing protein involves interaction with the mod(mdg4) protein (Gerasimova et al., 1995; Gdula et al., 1996; Georgiev and Kozycina, 1996; Gdula and Corces, 1997; Gerasimova and Corces, 1998). Thus, proteins that interact with, or bind to, CTCF are likely to
5 exist, although none are presently known.

According to the present invention, an insulator element, or CTCF binding site, is preferably located between an enhancer and a promoter to influence expression. The position of the insulator is the determining factor -- it can be inserted in either orientation with equal effect and insulator function. With regard to
10 current understanding of how enhancers function, various models have been proposed to account for enhancer blocking. The models fall into two broad categories: steric models and tracking models (Kellum and Elgin, 1998; Udvardy, 1999; Bell and Felsenfeld, 1999). Steric mechanisms postulate that insulators partition an enhancer and a promoter into two separate domains that are inaccessible
15 to each other. The steric models are related to existing ideas of how enhancers work. There is strong evidence that enhancers recruit the RNA polymerase complex to the promoter through interactions between proteins bound to that complex and proteins bound to the enhancer. If this occurs through formation of a loop between the enhancer site and the promoter, then the enhancer will be blocked if looping is
20 prevented. Tracking models presume that some activating signal must travel along the DNA from enhancer to promoter, and that the insulator blocks this transmission. Such activating signals might involve replication, or might for example, require that a polymerase complex travel along the DNA to reach the promoter. The identification of CTCF as a vertebrate enhancer blocking protein provides the ability
25 to functionally dissect the enhancer blocking process.

CTCF is likely to play a role in the function of many insulator elements. The first vertebrate insulator to be identified was located at the 5' end of the chicken β -globin domain; this 5' insulator site is likely to serve to protect the globin genes from inappropriate interaction with neighboring genes and their
30 regulatory elements (U.S. Patent No. 5,610,053 to J. Chung et al.). In particular, an independently regulated gene coding for a folate receptor has recently been identified 5' of the globin locus. The globin and folate receptor genes are close

– 15 –

enough so that the regulatory elements of the two loci might influence each other inappropriately in the absence of an insulator. A similar situation exists in the case of the T-cell receptor locus (Zhong and Krangel, 1997), where the BEAD insulator element shields against inappropriate activity of an enhancer. The presence of

5 CTCF DNA binding sites in these quite different genetic loci implies that the role of such sites in the establishment and maintenance of enhancer boundaries is likely to be a conserved and important component of gene regulation.

According to the present invention, the insulator element of SEQ ID NO:1 demonstrates enhancer-blocking function, both by itself and when bound by

10 the CTCF protein as described herein. Thus, this element and the CTCF protein can be regarded, in a broad sense, as a receptor and its ligand. These two entities can be used together or separately to regulate gene expression. The insulator defined herein is a DNA sequence which is capable of acting as a barrier to neighboring *cis*-acting elements, insulating the transcription of a gene placed within its range of

15 action, when juxtaposed between an enhancer and a promoter. Gene activation by external endogenous enhancers is blocked when the insulator is positioned between the enhancer and the promoter of a given gene.

A significant advantage of the insulator sequence defined by SEQ ID NO:1 is that it is a small molecule and is more versatile for use in a variety of

20 vectors for gene delivery into cells and organisms. By contrast, the larger 1.2 kb insulator and 250 bp core sequences are cumbersome and their sizes may preclude their use in some applications of gene delivery and/or gene transfer. Indeed, according to the results herein, the DNA motif which comprises the insulator of the present invention has been found to be both necessary and sufficient for insulating

25 and enhancer-blocking effects and so may be preferentially used as the insulator of choice in the vectors and constructs embraced by the invention.

Another aspect of the insulator sequence described herein, or the insulator bound by its cognate DNA binding protein, is the protection of a stably integrated reporter gene from position effects.

30 In one embodiment, the present invention provides constructs or vectors containing the insulator sequence of SEQ ID NO:1, enhancer and promoter sequences, and at least one heterologous gene sequence encoding a protein,

- 16 -

polypeptide, or peptide, wherein the insulator sequence is situated between the enhancer and promoter sequences, and upstream of the gene sequence. The construct or vector provides the vehicle for introducing the heterologous gene into a cell where it is integrated into the DNA and expressed and where its expression is insulated from the unwanted or adverse effects of *cis*-acting elements or sequences in surrounding chromatin. Insulated gene expression and regulation of gene expression by the use of the insulator of the present invention can be further regulated or controlled by endogenous CTCF, or a CTCF-like protein which binds to the DNA binding site specified by SEQ ID NO:1. Alternatively, a gene encoding CTCF, or a gene encoding a protein having DNA binding function like CTCF, can be used in a vector or construct that is co-introduced into a cell and expressed to more precisely control the expression of the introduced heterologous gene.

The vectors or constructs as used herein broadly encompass any recombinant DNA material that is capable of transferring DNA from one cell to another. The vector as described in the above embodiment can represent a mini-locus which can be integrated into a mammalian cell where it can replicate and function in a host cell type-restricted and copy number dependent manner, independent of the site of integration. Thus, the expression and production of the introduced gene is insulated from any effects exerted by neighboring genetic loci or chromatin following integration.

The insulator element as described herein can be employed to provide novel constructs for the efficient isolation and protection of genes and for the undisturbed production of a particular protein or other molecule encoded by a gene used in the constructs introduced into cells. The insulator element of the invention may also be used to insulate particular genes introduced and subsequently expressed in transgenic animals, such as fruit flies (e.g., *Drosophila melanogaster*), mice, rats, rodents, higher mammals and the like. Constructs containing the insulator element of the invention may be introduced into early fetal or embryonic cells for the production of transgenic animals containing the functional insulator element and reporter gene transcription unit. By insulating a gene or genes introduced into the transgenic animal, the expression of the gene(s) will be protected from negative or inappropriate regulatory influences in the chromatin at or near the

– 17 –

site of integration. In addition, the insulator will prevent inappropriate or unwanted activity from external enhancers that may affect the expression of the gene that has integrated into the DNA of a host cell.

The use of constructs harboring the insulator segment is envisioned
5 for the creation of knockout mice to determine the effects of a gene on development, or for the testing of therapeutic agents, such as chemotherapeutic or other types of drugs.

In general, the constructs of the present invention contain the
insulator sequence of SEQ ID NO:1, an enhancer element and a transcription unit
10 comprising, at a minimum, a gene of interest, for example, a gene encoding a protein or precursor thereof, and a promoter to drive the transcription of the gene of interest, and other sequences necessary or required for proper gene transcription and regulation (e.g. start and stop sites, origin of replication, splice sites and
polyadenylation signal). The enhancer is located in sufficient proximity to the
15 transcription unit to enhance the transcription thereof. The constructs may contain more than one small insulator of the invention, preferably in tandem, which are positioned so as to insulate the reporter gene and its transcription unit from surrounding DNA at the site of integration.

Transcriptionally competent transcription units can be made by
20 conventional techniques. In a preferred aspect of the present invention, the insulator element is situated between the enhancer and the promoter of a given gene to buffer the effects of a *cis*-acting DNA region on the promoter of the transcription unit. In some cases, the insulator can be placed distantly from the transcription unit. In addition, the optimal location of the insulator element can be determined by routine
25 experimentation for any particular DNA construct. The function of the insulator element is substantially independent of its orientation, and thus the insulator can function when placed in genomic or reverse genomic orientation with respect to the transcription unit to insulate the gene from the effects of *cis*-acting DNA sequences of chromatin.

30 The constructs as described herein may be used in gene transfer and gene therapy methods to allow the protected expression of one or more given genes that are stably transfected into the cellular DNA. The constructs of the invention

– 18 –

would not only insulate a transfected gene or genes from the influences of DNA surrounding the site of integration, but would also prevent the integrated constructs from impacting on the DNA at the site of integration and would therefore prevent activation of the transcription of genes that are harmful or detrimental to the cell.

5 The specificity of the constructs of the invention involves transfecting the particular gene(s) of interest into a cell type having the appropriate milieu for transcription of the gene(s) whose products are desired to be expressed. The constructs of the invention are capable of being transfected into a variety of cell and tissue types. In addition, since the insulator element itself is not cell or tissue
10 specific, it is a universal element which can act as a part of the constructs of the invention to insulation gene expression in the absence of strict cell or tissue specificity. The constructs can be designed to contain the appropriate regulatory sequences and all of the necessary DNA elements for integration of the construct and/or the appropriate components thereof and expression of a gene of interest in a
15 given cell type.

For assembly of the construct, the insulator element for ligation can be positioned in accordance with the desired use of the constructs of the invention. Thus, as disclosed above, at least one insulator may be positioned between an enhancer element and a promoter in a transcription unit, or the insulator can be
20 otherwise positioned on either side of a gene so as to obtain optimal insulation of the gene or genes desired to be transcribed. The insulator element can be obtained from natural sources or by synthetic means. For example, the insulator element can be excised from genomic or cDNA clones of eukaryotes, including chickens, mice, and humans, and the like, and then ligated with segments of DNA comprising the
25 enhancer and the transcription unit. Alternatively, the insulator element can be synthetically produced by conventional techniques of DNA synthesis such as the phosphite triester chemistry method (for example, see U.S. Patent No. 4,415,732 to Caruthers et al.; and Sinha, N.D. et al., 1984).

Those skilled in the art will appreciate that a variety of enhancers,
30 promoters, and genes are suitable for use in the constructs of the invention, and that the constructs will contain the necessary start, termination, and control sequences for proper transcription and processing of the gene of interest when the construct is

- 19 -

introduced into a vertebrate cell, such as that of mammal or a higher eukaryote. The constructs may be introduced into cells by a variety of gene transfer methods known to those skilled in the art, for example, gene transfection, lipofection, microinjection, electroporation, transduction and infection. In addition, it is
5 envisioned that the invention can encompass all or a portion of a viral sequence-containing vector, such as those described in U.S. Patent No. 5,112,767 to P. Roy-Burman and D.A. Spodick, for targeted delivery of genes to specific tissues. It is preferred that the constructs of the invention integrate stably into the genome of specific and targeted cell types.

10 Further, the DNA construct comprising the insulator element, enhancer and transcription unit may be inserted into or assembled within a vector such as a plasmid or virus, as mentioned above. The construct can be assembled or spliced into any suitable vector or cosmid for incorporation into the host cell of interest. The vectors may contain a bacterial origin of replication so that they can be
15 amplified in a bacterial host. The vectors may also contain, in addition to a selectable marker for selection of transfected cells, as in the exemplary constructs, another expressible and selectable or marker gene of interest.

Vectors can be constructed which have the insulator element in appropriate relation to an insertion region for receiving DNA encoding a protein or precursor thereof. The insertion region can contain at least one restriction enzyme
20 recognition site.

A particularly useful vector for gene therapy is the retroviral vector. A recombinant retroviral vector may contain the following parts: an intact 5' LTR from an appropriate retrovirus, such as MMTV, followed by DNA containing the
25 retroviral packaging signal sequence; the insulator element placed between an enhancer and the promoter of a transcription unit containing the gene to be introduced into a specific cell for replacement gene therapy; a selectable gene as described below; and a 3' LTR which contains a deletion in the viral enhancer region, or deletions in both the viral enhancer and promoter regions. The selectable
30 gene may or may not have a 5' promoter that is active in the packaging cell line, as well as in the transfected cell.

– 20 –

The recombinant retroviral vector DNA can be transfected into the amphotrophic packaging cell line Ψ -AM (see Cone, R. and Mulligan, R., 1984) or other packaging cell lines which are capable of producing high titer stocks of helper-free recombinant retroviruses. After transfection, the packaging cell line is
5 selected for resistance to G418, present at appropriate concentration in the growth medium.

Adenoviral vectors (e.g. DNA virus vectors), particularly replication-defective adenovirus vectors, or adeno-associated vectors, are also suitable for use and have been described in the art (S. Kochanek et al., 1996; G. Ascadi et al., 1994;
10 Ali et al., 1994).

Examples of transfectable reporter or heterologous genes that can be used in the present invention include those genes whose function is desired or needed to be expressed *in vivo* or *in vitro* in a given cell or tissue type. Genes having significance for genetic or acquired disorders are particularly appropriate for
15 use in the constructs and methods of the invention. Genes that may be insulated from *cis*-acting regulatory sequences by the insulator elements of the present invention may be selected from, but are not limited to, both structural and non-structural genes, or subunits thereof. Examples include genes which encode proteins and glycoproteins (e.g. factors, cytokines, lymphokines), enzymes (e.g. key
20 enzymes in biosynthetic pathways), hormones, which perform normal physiological, biochemical, and biosynthetic functions in cells and tissues. Other useable genes are selectable antibiotic resistance genes (e.g. the neomycin phosphotransferase gene (Neo[®]) or the methotrexate-resistant dihydrofolate reductase (dhfr) gene) or drug resistance genes (e.g. the multi-drug resistance (MDR) genes), and the like.

Further, the genes may encode a precursor of a particular protein, or the like, which is modified intracellularly after translation to yield the molecule of interest. Further examples of genes to be used in the invention may include, but are not limited to, erythroid cell-specific genes, B-lymphocyte-specific genes, T-lymphocyte-specific
25 genes, adenosine deaminase (ADA)-encoding genes, blood clotting factor-encoding genes, ion and transport channel-encoding genes, growth factor receptor- and hormone receptor-encoding genes, growth factor- and hormone-encoding genes, insulin-encoding genes, transcription factor-encoding genes, protooncogenes, cell
30

– 21 –

cycle-regulating genes, nuclear and cytoplasmic structure-encoding genes, and enzyme-encoding genes.

The present invention is also applicable to targeting tumor or malignant cells with the insulator element-containing constructs carrying genes
5 encoding toxins or toxoids, e.g. diphtheria toxoid and the like, to kill or otherwise damage and destroy the targeted cells. In addition, newly-cloned and isolated genes may be suitable candidates for use as reporter genes in the present invention.

Examples of eukaryotic promoters suitable for use in the invention are may include, but are not limited to, the thymidine kinase (TK) promoter, the
10 alpha globin, beta globin, and gamma globin promoters, the human or mouse metallothionein promoter, the SV40 promoter, retroviral promoters, cytomegalovirus (CMV) promoter, and the like. The promoter normally associated with a particular structural gene which encodes the protein of interest is often desirable, but is not mandatory. Accordingly, promoters may be autologous
15 (homologous) or heterologous. Suitable promoters may be inducible, allowing induction of the expression of a gene upon addition of the appropriate inducer, or they may be non-inducible.

Further, a variety of eukaryotic enhancer elements may be used in the constructs of the invention. Like the promoters, the enhancer elements may be
20 autologous or heterologous. Examples of suitable enhancers include, but are not limited to, erythroid-specific enhancers, (e.g. as described by Tuan, D. et al., and in U.S. Patent No. 5,126,260 to I.M. London et al.), the immunoglobulin enhancer, virus-specific enhancers, e.g. SV40 enhancers, or viral LTRs, pancreatic-specific enhancers, muscle-specific enhancers, fat cell-specific enhancers, liver specific
25 enhancers, and neuron-specific enhancers.

Many types of cells and cell lines (e.g. primary cell lines or established cell lines) and tissues are capable of being stably transfected by or receiving the constructs of the invention. Examples of cells that may be used include, but are not limited to, stem cells, B lymphocytes, T lymphocytes,
30 macrophages, other white blood lymphocytes (e.g. myelocytes, macrophages, monocytes), immune system cells of different developmental stages, erythroid lineage cells, pancreatic cells, lung cells, muscle cells, liver cells, fat cells, neuronal

- 22 -

cells, glial cells, other brain cells, transformed cells of various cell lineages corresponding to normal cell counterparts (e.g. K562, HEL, HL60, and MEL cells), and established or otherwise transformed cells lines derived from all of the foregoing. In addition, the constructs of the present invention may be transferred by various means directly into tissues, where they would stably integrate into the cells comprising the tissues. Further, the constructs containing the insulator elements of the invention can be introduced into primary cells at various stages of development, including the embryonic and fetal stages, so as to effect gene therapy at early stages of development.

10 In another embodiment of the invention, the constructs may be designed to contain genes encoding two subunits or components of a single protein so that each chain could be expressed from the same plasmid or suitable vector. For example, some proteins such as growth factors, growth factor receptors, blood clotting factors, and hormones are frequently comprised of two chains or subunits (e.g. α and β) which associate to form the functional molecule. In this embodiment, the gene coding for one chain or subunit of the molecule can be positioned in the plasmid or vector in conjunction with the insulator elements and specific promoter and enhancer elements (or heterologous promoter and enhancer, if desired), and the gene coding for the other chain or subunit can be positioned in the same plasmid or vector in conjunction with its insulator, promoter, and enhancer elements. The plasmid or vector containing the dual chain-encoding genes with their appropriately-positioned insulator elements can be transfected into cells to allow for the expression of a complete, two-chained molecule from the incorporated plasmid DNA, with each chain being regulated independently and with the copy numbers remaining the same.

25 When used in gene transfer and gene therapy, the constructs described herein may be administered in the form of a pharmaceutical preparation or composition containing a pharmaceutically acceptable carrier, diluent, or a physiological excipient, in which preparation the vector may be a viral vector construct, or the like, to target the cells, tissues, or organs of interest. The composition may be formed by dispersing the components in a suitable pharmaceutically-acceptable liquid or solution such as sterile physiological saline or

– 23 –

other injectable aqueous liquids. The composition may be administered parenterally, including subcutaneous, intravenous, intramuscular, or intrasternal routes of injection. Also contemplated are intranasal, peritoneal or intradermal routes of administration. For injectable administration, the composition is in sterile
5 solution or suspension or may be emulsified in pharmaceutically- and physiologically-acceptable aqueous or oleaginous vehicles, which may contain preservatives, stabilizers, and material for rendering the solution or suspension isotonic with body fluids (i.e. blood) of the recipient. Excipients suitable for use are water, phosphate buffered saline, pH 7.4, 0.15 M aqueous sodium chloride solution,
10 dextrose, glycerol, dilute ethanol, and the like, and mixtures thereof. The amounts or quantities, as well as routes of administration, used are determined on an individual basis, and correspond to the amounts used in similar types of applications or indications known to those of skill in the art.

Also contemplated by the invention is a kit or kits containing
15 insulator constructs in which the insulator elements of the invention are provided in a DNA receivable vector or plasmid that contains or can be readily adapted by the user to contain the appropriate DNA elements for proper expression of a gene or genes of interest. The insulator element-containing plasmids or vectors of the kit contain insulator elements, enhancers, a transcription unit, and the gene or genes of
20 interest may be inserted downstream of the insulator(s), as desired. Alternatively, the constructs of the kit may contain some or all of the necessary genetic elements for proper gene expression, or combinations of these, and the remaining genetic elements may be provided and readily inserted by the user, preferably between the insulator elements in the construct. The insulator element-containing plasmids or
25 vectors may be provided in containers (e.g. sealable test tubes and the like) in the kit and are provided in the appropriate storage buffer or medium for use and for stable, long-term storage. The medium may contain stabilizers and may require dilution by the user. Further, the constructs may be provided in a freeze-dried form and may require reconstitution in the appropriate buffer or medium prior to use.

30 The present invention further provides an insulator element found in the insulin-like growth factor 2 (Igf 2) locus. The insulator element contains a set of CTCF binding sites and comprises the sequence shown in SEQ ID NOS:84-87

– 24 –

(mouse), SEQ ID NOS:88-91 (rat), or SEQ ID NOS:92-98 (human). The *Igf2* gene encodes a growth promoting (or mitogenic) protein and the expression of this gene and its neighboring gene *H19* are imprinted. Expression of *Igf2* occurs exclusively from the paternally inherited allele. The insulator element identified is within a
5 region between the *Igf2* and *H19* genes that is methylated in the paternal allele only.

According to the present invention, the enhancer-blocking activity of the insulator element identified in the *Igf2* locus is dependent upon CTCF binding to the insulator. Methylation of the insulator element (cytosines in the sequence that are followed by a guanine (CpG) are methylated) abolishes the ability of CTCF to
10 bind to the insulator and would result in loss of CTCF-dependent enhancer-blocking activity. Therefore, on the paternal allele, methylation of the insulator element prevents CTCF binding to the insulator element which results in *Igf2* expression, whereas in the maternal allele where the insulator element is unmethylated, CTCF is capable of binding to the insulator element and prevent *Igf2* expression.

15 As methylation is likely to be the general mechanism by which an insulator whose enhancer-blocking activity is dependent upon CTCF binding is regulated, the invention also provides methods of modulating the CTCF-dependent enhancer-blocking activity of an insulator element by targeted methylation and demethylation of the insulator element. The gene of interest may be introduced into
20 cells by employing constructs or vectors in which the insulator element is strategically positioned with respect to the gene of interest, the promoter and the enhancer element so as to regulate the expression of the gene as described above. In one embodiment, a methylase is employed to methylate the insulator element, thereby activating the expression of the gene of interest in the cell. In one method,
25 DNA methyltransferase 3, which has been shown to be capable of de novo methylation of the cytosine of the CpG residues both in vivo and in vitro (See Bird, 1999 for review), is employed. In this method, in addition to the gene of interest, the promoter, the enhancer element and the insulator element, the vector introduced into the cells also comprises a DNA binding sequence of a DNA binding protein, for
30 example, Gal4 or LexA. The DNA binding sequence would be located in regions adjacent to the CpG residues to be methylated in the insulator element. A vector encoding a fusion protein in which the enzymatic domain of the DNA

– 25 –

methyltransferase 3 is fused to the DNA binding domain of the DNA binding protein is then introduced into the cells. Methods for generating such domain-fusion proteins are well known to those skilled in the art. Binding of the fusion protein to the DNA binding sequence adjacent to the CpG residues and the
5 expression of the fusion protein allow the CpG residues to be methylated by the methyltransferase. In another embodiment, a similar method may be employed to prevent the expression of the gene of interest in a cell by demethylating the CpG residues using a demethylase.

Examples

10 The examples as set forth herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the invention in any way.

Example 1

Materials and Methods

15 A. Plasmid Construction and Oligonucleotides

The plasmid pNI was the base plasmid in which all DNA fragments were tested for enhancer blocking activity.

DNA fragments subcloned into the AscI site of this plasmid are located between the enhancer (mouse HS2) and the reporter ("γ-neo"). When
20 cloned into the NdeI site of pNI, inserted DNA sequences are located "upstream" of the enhancer, and when cloned into the XbaI site, the sequences are located "downstream" of the reporter; in both of these cases the insert is located outside the promoter-enhancer path.

pNI was generated by replacing the SacI copy of the 1.2 kb insulator
25 (found between the enhancer and the promoter) in pJC5-4 (Chung et al, 1993) with an AscI linker (New England Biolabs) after digestion of this plasmid with EcoII36II. The following primers were used in PCR amplifications to generate fragments for cloning into the AscI site of pNI:

30 AC1F (SEQ ID NO:29); AC1R (SEQ ID NO:30); AC2F (SEQ ID NO:31); AC2R (SEQ ID NO:32); ΔIACF (SEQ ID NO:33); ΔIacF (SEQ ID NO:34); ΔIacR (SEQ ID NO:35); ΔIIIACF (SEQ ID NO:36); ΔIIIACR (SEQ ID

– 26 –

NO:37); Δ IVACF (SEQ ID NO:38); Δ IVACR (SEQ ID NO:39); Δ VACR (SEQ ID NO:40); BEADascF (SEQ ID NO:41); BEADascR (SEQ ID NO:42); BEAD Δ AF (SEQ ID NO:43); and BEAD Δ AR (SEQ ID NO:44).

The core, FI/FII, FIII/FIV/FV, Δ I and Δ V were generated by PCR
 5 using the plasmid p50I (Reitman and Felsenfeld, 1990) as a template and the primer pairs AC1F/AC2R, AC1F/AC1R, AC2F/AC2R, Δ IACF/AC2R, and AC1F/ Δ VACR, respectively. Deletions of FII, FIII and FIV from the core were accomplished by two-step, overlapping PCR. For each deletion, a pair of intermediate fragments was generated by PCR in separate reactions using p50I as the template and the primer
 10 pairs AC1F/ Δ IacR, Δ IacF/AC2R, AC1F/ Δ IIACR, Δ IIACF/AC2R, AC1F/ Δ IVACR, and Δ IVACF/AC2R. The products of each of these reactions were gel-purified, mixed pair-wise to generate the appropriate templates, and the final products were amplified with AC1F/AC2R (for example, the products of amplifications with AC1F/ Δ IacR and Δ IacF/AC2R were mixed to generate the
 15 template for PCR of the core- Δ II).

The full-length BEAD-1 fragment was generated by PCR from K562 genomic DNA with primers BEADascF and BEADascR. The fragment BEAD Δ A was generated in a two-step, overlapping PCR reaction, first using the BEAD-1 fragment as a template and the primers BEADascF/BEAD Δ AR and BEAD Δ AF/
 20 BEADascR in separate reactions, then mixing the gel-purified products of these reactions with primers BEADascF and BEADascR to generate the final product by PCR.

The a 1.6 kb fragment containing the full length RO element was subcloned into Ec1136II cut pJC5-4 after liberation of this fragment from p0, 1
 25 (Robinett et al., 1997) by digestion with Ec1136II and PvuII. All other enhancer-blocking fragments were generated by direct synthesis of the appropriate complementary oligonucleotides on an ABI 394 DNA synthesizer. The top strands of these were FII/FIII: (SEQ ID NO:45); FII/III- Δ Sp1* (SEQ ID NO:46); FII/III- Δ α 2 (SEQ ID NO:47); Δ spacer (SEQ ID NO:48); Δ IIN (SEQ ID NO:49); Δ IIIN
 30 (SEQ ID NO:50); FII (SEQ ID NO:51); FIII (SEQ ID NO:52); gypsy-3 (SEQ ID

– 27 –

NO:53); mycFV (SEQ ID NO:54); lys (SEQ ID NO:55); ApB (SEQ ID NO:56); RO100 (SEQ ID NO:57); and BEAD-A (SEQ ID NO:58).

All FII mutants were identical to FII, except for those bases indicated in lowercase in Fig. 3D. For use in the enhancer-blocking assay, single stranded
 5 oligonucleotides were purified by denaturing PAGE, quantified, annealed, digested with AscI, and subcloned into pNI. The FII site was also generated as above with NdeI sites at its ends for cloning upstream of the enhancer in pNI to generate FII-UP. To generate FII-DOWN, FII was digested out of pNI-FII, the ends were flushed with Klenow, and XbaI linkers (New England Biolabs) were added for
 10 cloning into the XbaI site of pNI.

The 2.5 kb Imprinted Control Region (ICR) fragment and the 1.6 kb deleted fragment (DMD) within the ICR were generated by PCR on genomic DNA with ICRR (SEQ ID NO:62) and ICRF (SEQ ID NO:63); ICRR and DMDF (SEQ ID NO:64) primers, respectively. The ~800 bp HS1 fragment was generated with
 15 ICRF and HS1R (SEQ ID NO:65); and the HS2 fragment was generated with HS2F (SEQ ID NO:66) and ICRR. Deletions of m1 and m2 from HS1 were accomplished by PCR using the following additional primers: HS1Δm1F (SEQ ID NO:67); HS1Δm1R (SEQ ID NO:68); HS1Δm2F (SEQ ID NO:69); HS1Δm2R (SEQ ID NO:70); HS2Tm4R (SEQ ID NO:71); HS2F (SEQ ID NO:72); HS2Tm3F (SEQ ID
 20 NO:73); HS2Δm3F (SEQ ID NO:74); HS2Δm3R (SEQ ID NO:75). The fragments Δm3 and Δm4 are ~200 base pair truncations of the 5' and 3' of HS2 generated by PCR with the primer pairs HS2Tm3F/ICRR and HS2F/HS2Tm4R, respectively. In the fragment Δm3Δm4, ~90 base pairs spanning the m3 site were internally deleted while the deletion of m4 results from a 3'-truncation. This was accomplished by
 25 two-step overlapping PCR using the primer pairs HSΔm3F/HS2Tm4R and HS2F/HS2Δm3R on a DMD clone template. The products of these reactions were gel-purified, mixed, and the final product was amplified by PCR with the primer pairs HS2F/HS2Tm4R. Internal deletions of ~90 bp fragments spanning m1 and m2 from HS1 were generated by first amplifying with the primer pairs
 30 HS1F/HS1Δm1R, HS1Δm1F/HS1R, HS1Δm1F/HS1Δm2R, HS1F/HS1Δm2R, and HS1Δm2F/HS1R. To generate singly or doubly deleted fragments the products of

- 28 -

these reactions were gel-purified, mixed accordingly, and amplified with HS1F/HS1R. The resulting fragments were sub-cloned into pNI after addition of the appropriate linkers where necessary. For enhancer blocking assays with m3, 5'-GCTGTTATGTGCAACAAGGGAACGGATGCTACCGCGCGGTGGCAGCATA
5 CTCCTATATATCGTGGCCCAAATGCTGCCAACTTGGGGGGAGCGATTCA
TTC (SEQ ID NO:83) was directly synthesized with the appropriate restriction sites at its ends and cloned into pNI at either the AscI or NdeI sites.

B. Enhancer Blocking Assay

Enhancer blocking assays were performed as previously described
10 (Chung et al., 1993 and Chung et al., 1997). Briefly, 20 µg of each construct was linearized by SalI digestion, phenol-chloroform extracted, ethanol precipitated, and quantified by UV absorption. 20 ng of each DNA was then electroporated into K562 cells (1×10^7) and after allowing 24 hours for recovery, cells were plated in soft agar with geneticin (Life Technologies) at 750 µg/ml (active). Colonies were
15 counted after 3 weeks of selection and the colony number was normalized to that obtained with pNI or a construct which had 2.3 kb of λ DNA inserted between the enhancer and the reporter as a spacer control.

C. DNase I Hypersensitive Site Analysis

Nuclei were isolated from adult chicken red blood cells, essentially
20 as described previously (Bresnick and Felsenfeld, 1994), except that 0.2 mM EGTA was included in all buffers. After incubation of the nuclei with varying concentrations of DNase I for 5 minutes at room temperature, the reaction was terminated by the addition of SDS and the genomic DNA was purified. To map precisely the position of HS4, DNase I digested and undigested genomic DNAs (10
25 µg) were further digested with StyI to generate an ~1 kb parent fragment which spanned the insulator core. StyI digested control DNAs were also digested with the enzymes indicated in Fig. 1A. All of these DNAs were resolved on a 1.3 % agarose gel and subjected to Southern blotting by standard techniques using a 503 bp StyI-SacI fragment from the plasmid p501 as a probe (Reitman and Felsenfeld, 1990).

30 D. DNA Binding Assays

– 29 –

All DNA binding assays were carried out in a binding buffer containing 20 mM HEPES, pH 7.9; 150 mM KCl; 5 mM MgCl₂; and 1 mM DTT. For gel mobility shift assays, DNA binding was carried out at room temperature for 30 minutes in binding buffer plus 5% glycerol, 20-40 fmol of labeled double stranded oligonucleotide probe, poly dI/dC at 50-100 µg/ml, and 1-5 µl of protein in a final volume of 20 µl. Probes were oligonucleotide duplexes identical to those used for subcloning into the enhancer-blocking assay. 10 pmol of each top strand was end-labeled with ³²P and then annealed with 15 pmol of an unlabeled complementary oligonucleotide; the resulting duplexes were used directly as probes. Cold competitor duplexes were added simultaneously with labeled probes at 50-fold molar excess.

An antibody was raised against a C-terminal peptide (APNGDLTPMILSMMD), SEQ ID NO: 59, of CTCF. Supershifts were carried out by pre-incubating the appropriate proteins with purified antibodies in binding buffer for 2 hours at 0° C, followed by a room temperature incubation of 30 minutes in the presence of DNA.

For southwestern assays, proteins were resolved by SDS-PAGE, transferred to PVDF, and then denatured and renatured by successive 10 minute incubations in binding buffer supplemented with guanidine hydrochloride at 4.8, 3, 1.5, and 0.75 M. After an additional 10 minute wash in binding buffer, the blots were blocked in binding buffer plus 5% non-fat dry milk for 16 hours at 4° C. An FII probe was generated for southwestern assays by annealing a full length top strand sequence: (CCCAGGGATGTAATTACGTCCCTCCCCCGCTAGGGGG-CAGCAGGCGCGCCT), (SEQ ID NO:60) to a short 3' complementary primer (AGGCGCGCCTGCTGC), (SEQ ID NO:61). This partial duplex was then extended with Klenow in the presence of α-³²P-dCTP resulting in a probe identical to that used in gel-shift assays but with 10 labeled phosphates per molecule. Blots were probed for 3 hours at room temperature in binding buffer supplemented with 0.25% non-fat dry milk, 5 µg/ml poly dI/dC and 3 pmol of labeled probe in a final volume of 20 ml, washed three times for 10 minutes in the same buffer without DNA and exposed to film.

– 30 –

Probes used in the DNA binding assays described in Example 6 were annealed duplexes of the following sequences: m1 (SEQ ID NO:76); m2 (SEQ ID NO:77); m3 (SEQ ID NO:78); m4 (SEQ ID NO:79); h1 (SEQ ID NO:80); FIIx3' (SEQ ID NO:81); m1x3' (SEQ ID NO:82).

5 E. Protein Purification and Translation

Nuclear extracts from K562 cells and whole chicken blood were prepared essentially as previously described (Evans et al., 1988). For purification of the FII binding protein, nuclei were prepared from 6 liters of whole chicken blood (Pelfreez Biologicals) and extracted in Buffer C: 20 mM HEPES, pH 7.9, 420 mM
10 NaCl, 5 mM MgCl₂, 0.2 mM EDTA and 1 mM DTT. The resulting extract was diluted to 150 mM NaCl and 20% glycerol and fractionated on a 500 ml SP sepharose column (Pharmacia) using a 0.15-1M NaCl linear gradient. Active fractions were pooled, diluted to 150 mM NaCl and loaded onto a 25 ml CM sepharose column (Pharmacia). Active fractions eluted with a peak at 600 mM
15 NaCl from a 0.15-1M NaCl gradient. These fractions were pooled and loaded onto a 2.6/60 cm Sephacryl S-300 gel filtration column (Pharmacia) that was pre-equilibrated with 20 mM HEPES, pH 7.9, 150 mM NaCl, 5 mM MgCl₂, 0.2 mM EDTA and 1 mM DTT. Active fractions were pooled, dialyzed into 10 mM sodium phosphate pH 8.0, 150 mM NaCl, 5 mM MgCl₂, 1 mM DTT and 20% glycerol, and
20 loaded onto a 25 ml Macro-Prep ceramic hydroxyapatite column (Bio-Rad). This column was eluted with a 10-800 mM phosphate gradient at pH 8.0. Throughout the isolation all buffers were supplemented with 1 mM PMSF, 0.7 µg/ml pepstatin, and 0.5 µg/ml leupeptin and maintained at 4° C. Fractions pooled from the gel-filtration, and all subsequent buffers, were also supplemented with 40 µg/ml bestatin
25 and AEBSF (at 200 µg/ml) was substituted for PMSF. Final active fractions were identified by gel shift assay, and analyzed by southwestern with an FII probe. For peptide sequencing, 1 ml of a final active fraction (representing ~1/10th of the final yield and ~5 µg of purified ~140 kDa protein) was TCA precipitated, resolved on
30 7% Tris-acetate SDS-PAGE (Novex), transferred to PVDF, stained with imido black, and internal protein sequence data were obtained at the Rockefeller University Protein/DNA Technology Center.

- 31 -

In vitro-translated human CTCF was obtained using the plasmid p4B7.1 as a template for *in vitro* transcription by T7 polymerase according to the manufacturers instructions (Ambion, 'Message Machine'), followed by *in vitro* translation of the resulting RNA in a nuclease treated rabbit reticulocyte system (Promega). The plasmid p4B7.1 contains the full length human CTCF cDNA (Filippova, et al., 1996) subcloned into pCITE4b(+) (Novagen).

Example 2

Identification of an Enhancer-Blocking DNA Fragment

A 1.2 kb DNA element at the 5' end of the chicken β -globin locus, corresponding to a constitutive DNase I hypersensitive site (5'HS4), was shown to function as an insulator in an enhancer-blocking assay (U.S. Patent No. 5,610,053 to J. Chung et al.; Chung et al., 1997). The assay tested the ability of a DNA sequence to prevent activation of a gene for neomycin resistance by a strong enhancer when the construct was stably transformed into an erythroleukemia cell line (Chung et al., 1993). The insulator effect was manifested by a marked reduction in the number of colonies resistant to G418 only when the globin sequence element was placed between the enhancer and the promoter. This same assay was utilized to show that a large part of the insulator activity was contained in a 250 bp GC-rich 'core' fragment at the 5' end of the 1.2 kb element. HS4 mapped precisely within this core region, consistent with its significance *in vivo*. (Fig. 1A).

DNase I footprinting of the 250 bp core insulator sequence with nuclear extracts revealed five protected regions (FI to FV), as illustrated in Fig. 1C; Chung et al. 1997). The FI-FV DNA segment was further dissected and analyzed to identify an insulator-protein binding site. The core region was divided into separate fragments; each fragment was employed in enhancer blocking assays. Splitting the core between FII and FIII generated two fragments (FI/FII and FIII/IV/V) each of which had some enhancer blocking activity. However a fragment containing only FII and FIII had greater activity than the entire core (FII/III, Fig. 1D).

Deletion analysis confirmed that FII and FIII were responsible for the majority of the enhancer blocking activity. While deletion of FI had a slight effect, deletion of FII and FIII significantly reduced enhancer-blocking activity (Fig. 1E).

– 32 –

Deletions of FIV and FV were essentially neutral. Considered together, these results show that regions FII and FIII represent a functional enhancer-blocking region of the core. Consistent with this conclusion, the insertion of an increasing number of copies of FII/FIII between the enhancer and the promoter resulted in a
5 linear increase in blocking activity, as was also observed for the 1.2 kb insulator and the 250 bp core sequence (Fig. 1F).

Further analysis of the FII/FIII element revealed an internal 'spacer' sequence that appeared to partially counteract the enhancer-blocking activity. Removal of this spacer region (Fig. 2A) resulted in even stronger blocking activity.
10 In fact, the removal of sequences adjacent to FII resulted in the discovery of an approximately 50 bp sequence spanning FII that alone was found to possess a blocking activity nearly equal to that of the full 1.2 kb insulator element. Consistent with the behavior *in vivo*, it is noted that the position of FII is coincident with that of HS4 in nuclei (Fig. 1A). Importantly, enhancer-blocking by these minimal
15 fragments, including FII, displayed the same position-dependence as that observed for the entire 1.2 kb insulator element. When placed either upstream of the enhancer or downstream of the promoter in the enhancer blocking assay, FII had essentially no effect on expression (Fig. 2B; Chung et al., 1997). Thus, in order to effect expression, FII must be located between the enhancer and the promoter (Fig.
20 2B).

Example 3

Identification of a Specific Enhancer-Blocking Protein

Experiments were conducted to identify proteins that bound to the FII fragment, in view of the strong enhancer-blocking activity exhibited by this
25 fragment. A comparison of the sequence of FII with that of known transcription factor binding sites revealed several potentially significant homologies (Fig. 2E). An Sp1 consensus sequence lies in the middle of the FII fragment and a sequence homologous to a yeast $\alpha 2$ binding site (Sauer et al., 1988) overlaps a partial match to the binding site of the *Drosophila* protein suppressor of Hairy-wing (Su(Hw)),
30 (Fig. 2E; Geyer and Corces, 1992). To test whether any of these homologies could account for the blocking activity of FII, mutations were introduced that were

– 33 –

predicted to reduce dramatically the affinity of each of these proteins for the sequence of the FII fragment. The mutations were introduced into either the FII or the FII/III fragments and were tested in the enhancer-blocking assay. A deletion of 4 base pairs within the region that overlaps both the $\alpha 2$ and the Su(Hw) binding sites had no effect on the blocking activity of the FII/III fragment (Fig. 2D). Furthermore a 100 bp fragment, derived from the *Drosophila gypsy* element and containing three canonical Su(Hw) binding sites had no activity in the assay. Thus, it was concluded that neither the Su(Hw) site nor the $\alpha 2$ site can account for the activity contained in FII.

Similarly, in the context of FII, mutation of the Sp1 consensus sequence had no effect on the blocking activity of the fragment; in fact, mutation of each of the three potential Sp1 binding sites in FII/III resulted in substantially increased activity (Fig. 2D). Sp1 may act as an inhibitor of enhancer-blocking in the enhancer-blocking assays described here. This may also explain the above-mentioned inhibitory effect of the "spacer" sequence between FII and FIII and account for the observation that the activities of FII and FIII are not always additive (see, for example, Δ spacer in Fig. 2A). In addition, mutation of the Sp1 sites in FII/III rendered the level of enhancer-blocking activity equal to the sum of that of FII and FIII.

To determine which sequences within FII were responsible for its activity, multiple transversions (e.g., C \rightarrow A and G \rightarrow T) were made across the 5', middle and 3' regions of the fragment (Fig. 3A). All of these transversions reduced the level of enhancer-blocking activity of FII, but changes at the 3' end of the fragment (x3') caused a complete loss of activity. In addition, deletion of 10 bp from both ends of FII (Δ F), or a reversal of the sequence 5' - 3' (rev), resulted in dramatic reductions in activity. An effect of sequence composition is ruled out by the "rev" mutant since its base composition is identical to that of FII.

In light of the foregoing, experiments were carried out to identify protein(s) that binds to FII with a competition profile that matched the sequence specificity observed in the enhancer-blocking assay. Nuclear extracts were prepared from the human erythroleukemic cell line K562 (the cell line in which the enhancer-

- 34 -

blocking assay was performed) and from adult chicken red blood cells (the FII insulator is a chicken-derived element). Identical patterns were obtained with these two extracts in a gel mobility-shift assay (Figs. 3B and 3C). For each cell source of nuclear extract, two major complexes were observed when the extract was incubated with a 60 bp probe spanning footprint II. The upper complex could be super-shifted with an antibody against Sp1 and was competed by a 100-fold excess of an unlabeled oligomer carrying an Sp1 consensus binding site. This complex was concluded to contain Sp1. In contrast, the lower complex was neither super-shifted by anti-Sp1 antibody, nor was its binding influenced by an excess of Sp1 consensus binding site. Importantly, the degree to which each of the tested fragments competed for binding to this complex paralleled its ability to act as an insulator in the enhancer-blocking assay (Fig. 3D).

Example 4

Isolation of a Protein Responsible for the Sequence Specificity of Enhancer-Blocking

Probing a southwestern blot of nuclear proteins with labeled FII revealed a single FII-specific DNA binding protein with an apparent size on gels of ~140 kDa (Fig. 4A). This protein was purified by conventional chromatography. Throughout the purification, the elution profiles of FII binding activities were identical in gel-mobility shift and southwestern assays. This protein bound tightly to S, CM, and hydroxyapatite columns, and eluted with a peak at ~330 kDa on gel filtration (Fig. 4B). Coomassie staining of gels of the final hydroxyapatite fractions revealed a single protein with an apparent molecular weight of 140 kDa corresponding to the position of the FII southwestern activity (Fig. 4D). The sequences of four internal peptides (Fig. 4C) from the 140 kDa DNA binding component of the final purified fraction all perfectly matched the predicted sequence of a previously cloned 11 zinc finger DNA binding protein, CTCF (Klenova, 1993; Filippova et al., 1996).

Consistent with this identification, *in vitro*-translated CTCF binds to FII with a sequence specificity identical to that observed in the gel-mobility shift and enhancer-blocking assays (Fig. 5). As expected, this protein also bound to other previously characterized CTCF sites (Fig. 5, lanes 9-11) and these sites also act as

- 35 -

enhancer blockers in our assay (Figure 5). Alignment of these CTCF sites with FII revealed a conserved region which has been shown to be critical for binding of CTCF to these other sites (Filippova et al., 1996; Burcin et al., 1997; Vostrov and Quitschke, 1997). Mutation of this conserved 3' sequence completely abrogated binding and enhancer blocking in the relevant binding and enhancer-blocking assays (see x3' in Figs. 3A-3D and alignment in Fig. 6A).

Example 5

Conservation of Sequence Among Vertebrate Insulators

Because CTCF is highly conserved among vertebrates, an investigation as to whether CTCF sites might be present in other vertebrate insulator elements was carried out. Two such elements have recently been described. A 1.4 kb fragment found in the intergenic spacer region of the ribosomal RNA genes of *Xenopus laevis*, termed the repeat organizer (RO), has been shown to prevent enhancer action in a directional manner (Robinett et al., 1997). The 3' half of this sequence is composed of seven tandem repeats of an ~100 bp GC-rich sequence (Labhart and Reeder, 1987). The RO sequences bear significant homology with CTCF sites, including FII (Fig. 6B). In the enhancer-blocking assay, the full-length RO conferred moderate enhancer-blocking activity and a single copy of the 100 bp RO repeat had weak enhancer blocking activity on its own (Fig. 7B). It is perhaps because of the weak activity of a single copy of this sequence, that attempts to obtain reproducible binding of CTCF to a single RO repeat have been unsuccessful.

The only other vertebrate insulator described to date, BEAD-1, is a 1.6 kb enhancer-blocking element derived from the human T-cell alpha/delta (α/δ) locus (Zhong and Krangel, 1997). Best-fit alignment of this element with various CTCF sites revealed a good match between FII and a-sequence roughly at the center of this element (BEAD-A in Fig. 6C). In fact, a DNA fragment containing this region also bound specifically to purified chicken CTCF (Fig. 7A, lanes 6 and 5 respectively). Consistent with these observations, both full-length BEAD-1 and the CTCF binding BEAD-A element defined here were effective enhancer-blocking elements in the described enhancer-blocking assay (Fig. 7B). Furthermore, deletion

– 36 –

of the BEAD-A sequence from BEAD-1 largely eliminated the activity of the larger element.

Example 6

Methylation of CTCF Binding Sites Controls Imprinted Expression of the *Igf2* gene

- 5 The gene encoding Insulin-like Growth Factor 2 (*Igf2*) and the *H19* gene are neighboring genes on the same chromosome. The *Igf2* gene encodes a growth promoting (or mitogenic) protein and abnormal expression of this gene has been linked to numerous cancers and is thought to play a causal role in the etiology of several growth defect-related syndromes.
- 10 Expression of the *Igf2* and *H19* genes is imprinted. Genomic imprinting refers to a mechanism through which expression of a particular gene is dependent upon the gamete (or parent) of origin. For example, a gene is expressed when on the chromosome contributed by the mother, but not expressed from the paternal chromosome. As the alleles are identical in sequence, the signal for such a
- 15 mechanism cannot rely on DNA sequence. Gamete specific modification of the DNA though DNA methylation is believed to play a major role in specifying the allele's parent-of-origin.
- Although the *Igf2* and *H19* genes share an enhancer (Yoo-Warren, 1988), *H19* is only expressed from the maternal allele, while expression of *Igf2*
- 20 occurs exclusively from the paternally inherited allele (Bartolomer, 1991; Dechiara, 1991). A region located upstream of the mouse *H19* gene and between the *Igf2* and *H19* genes is methylated in the paternal allele only. This differentially methylated region appears to be the site of an epigenetic mark that is required for the imprinting of these genes. One study has shown that a deletion within this region results in loss
- 25 of imprinting of both *H19* and *Igf2* and maternal transmission of a 1.6 kb deletion within this region results in expression of the normally silent *Igf2* allele (Thorvaldsen, 1998). The ability of this deleted fragment (DMD) to act as a positional enhancer-blocking element was examined by inserting it at various locations relative to an enhancer as shown in Figure 8. Insertion of the 1.6 kb DMD
- 30 fragment between the enhancer and the promoter results in an 8–10-fold drop in colony number, similar to the 8-fold drop observed with the previously

- 37 -

characterized 1.2 kb chicken β -globin insulator (U.S. Patent No. 5,610,053). These results cannot be explained by the increased distance between the enhancer and the promoter, as insertion of up to 2.3 kb of heterologous DNA between them has little effect on colony number (U.S. Patent No. 5,610,053). Furthermore, like the β -globin insulator, when the DMD is placed outside the enhancer-promoter path, either upstream of the enhancer or downstream of the promoter, it has little effect on expression (Fig. 8b). Therefore, DMD appears to have the position-dependent enhancer-blocking properties of an insulator.

The DMD fragment is part of a slightly larger (~2 kb) imprinted control region (ICR) that is methylated throughout development exclusively on the paternal allele (Tremblay, 1995 and 1997). Allele-specific alterations in chromatin structure were also observed in this region (Szabo, 1998; Khosla, 1999 and Hark, 1998). Two nuclease-hypersensitive regions are located exclusively on the maternal allele (HS1 and HS2 in Fig. 8a), whereas the chromatin on the paternally derived allele is methylated and nuclease insensitive (Hark, 1998). Both HS1 and HS2 remain hypersensitive throughout development and are present independent of tissue type. The enhancer-blocking potential of fragments spanning HS1, HS2, and a larger fragment that spans the entire ICR was tested. All of these fragments confer enhancer-blocking activity (Fig. 8b). HS1 and HS2 individually show considerable enhancer-blocking activity; a fragment that contains both HS1 and HS2 essentially eliminates the enhancer's influence on expression (Fig. 8b, compare NI E with ICR).

A BestFit comparison between the FII fragment of β -globin and the 2.6 kb of sequence spanning the ICR revealed a 13/16 match between the 3'-end of FII and a sequence at the 5'-edge of HS2 (m3 in Fig. 9a). By searching the remainder of the ICR with the m3 sequence, a total of four homologous sequences (m1-4 in Fig 9a) was identified. Consistent with their *in vivo* significance, sequences homologous to these mouse sites are also found upstream of the human and rat *H19* genes (aligned in Fig. 9a); conservation of sequences overlapping those shown here was recently noted (Frevel, 1999a; Stadnick, 1999). In humans, a region of paternal-specific DNA methylation upstream of *H19* has also been defined (Jinno, 1996; Frevel, 1999b) and these homologous sequences are part of a larger

- 38 -

repeating element found in that region. An alignment of all of the sites from rat, human, and mouse reveals a 12 base pair consensus sequence that is shared among them (Fig. 9a). This consensus bears an 11/12 match with the sequence at the 3' of *β-globin* FII. In addition, each of these mouse sites, and a representative human site
5 bind to both purified chicken CTCF (P) and *in vitro* translated human CTCF (I) (Figure 9c). Like FII, a fragment spanning a single mouse ICR site confers position-dependent enhancer blocking activity *in vivo* (Fig. 9b). Furthermore, in gel shifts with K562 nuclear extracts (E) (the cells in which the enhancer blocking assays were performed) a complex that comigrates with the FII/CTCF complex was
10 observed with each of these mouse and human ICR sites (Figure 9c) and an antibody raised against CTCF supershifts this complex (Fig. 10d). Consistent with the enhancer blocking activity of CTCF, this complex was competed by FII, but not by a mutant of FII in which both enhancer blocking and CTCF binding have been eliminated (FIIx3' in Fig. 9d). When the base pairs shared among the ICR sites and
15 FII are altered in the context of one of these mouse sites it no longer competes for binding to CTCF (m1x3' in Fig. 9d).

In the mouse *H19* ICR, HS1 and HS2 each contain two CTCF sites. The two sites were deleted sequentially and the enhancer blocking activities of the resulting fragments were measured (Fig. 10). Since the enhancer blocking activities
20 of HS1 and HS2 are somewhat dependent upon their orientations (data not shown), deletion analyses were carried out with the orientation that gave the strongest activities. In each case, a deletion that eliminates either one of the CTCF sites results in a reduction in enhancer blocking activity, while deletion of both sites from either HS1 or HS2 eliminates their activity (Fig. 10a). The deletions span sequence
25 that are larger than the average ~53 bp CTCF footprint. Among these sequences, however, the only significant similarity is within the CTCF sites. These similarities define a consensus for CTCF binding (Fig. 9a and d) which is essential to the enhancer blocking activity of *β-globin* FII. As shown in Fig. 9b, single CTCF site from several other loci (including the mouse ICR0 alone confer enhancer blocking
30 activity.

The above results demonstrate that sequences within the mouse *H19*, ICR have the enhancer blocking properties of an insulator. Several recent studies

– 39 –

suggest that this activity is directly involved in the regulation of *Igf2*. One study showed that if the *H19* enhancer is moved from its genomic location downstream of *H19* to a new location upstream of the ICR, the normally silent maternal allele of *Igf2* is expressed (Webber, 1998). This suggests that it is the enhancer's position downstream of the *H19* locus that prevents activation of the maternal *Igf2* allele. Competition between the *H19* and *Igf2* promoters cannot explain this result since deletion of the *H19* promoter has no effect on *Igf2* expression (Schmidt, 1999). Instead, it is the enhancer's position relative to the ICR that restricts its action: a deletion within the ICR results in biallelic expression of *Igf2* (Thorvaldsen, 1998). This line of reasoning is further supported by the observation that maternal inheritance of the relocated enhancer results in loss of expression of the normally active *H19* allele, in this case because the ICR, now located between the enhancer and the *H19* promoter, blocks their interaction. Thus, the dependence of *H19* and *Igf2* expression on the position of the *H19* enhancer is explained by a single model that posits the existence of an insulator within the ICR (Thorvaldsen, 1999; Webber, 1998; Leighton, 1995).

As the *H19* locus contains an insulator that is active only on the unmethylated (maternal) allele, a model has been proposed which suggests that the influence of the ICR on expression of *Igf2* depends upon the allele's parent of origin (Leighton, 1995). In this model, inheritance of paternal-specific CpG methylation in the ICR results in inactivation of the insulator and thus on this allele the *H19* enhancer is free to activate *Igf2*. Direct support for a role of DNA methylation in activation of *Igf2* comes from the observation that in DNA methyltransferase-1 deficient mouse embryos, both alleles of *Igf2* are silent (Li, 1993).

The results of the deletion analysis of HS1 and HS2 imply that it is the conserved CTCF sites in these elements that are responsible for their enhancer blocking activities. One model that could explain why CpG methylation abolishes this activity is that CTCF cannot bind these sites when they are methylated. To test this, the corresponding oligomers were synthesized with ⁵meC incorporated at each CpG and the ability of the resulting duplex to compete for the binding of CTCF to the unmethylated form was assessed. Methylation of each of the mouse sites, and a representative human site, greatly reduces their ability to compete for binding of

– 40 –

CTCF to an unmethylated site, even at a 50-fold molar excess (Fig. 10b).

Methylation of β -globin FII has a similar effect. Because FII and the ICR sites have only one CpG in common, the influence of methylation at only this site (on both strands) in several ICR sites was examined (Fig. 10b, right panel). In fact,

5 methylation of this CpG alone significantly reduced CTCF binding to all of these sites (Fig. 10b, M1 lanes). This result implies that enhancer access could, in principle, be regulated by a single (perhaps targeted) methylation event. Further examination of the influence of methylation on gene expression will require a system that allows for the establishment and maintenance of partially methylated
10 transgenes *in vivo*.

The above results demonstrate that the *H19* ICR is an enhancer-blocking element. CTCF binding sites are required for this activity and when these sites are methylated, they no longer bind the insulator protein CTCF. These results provide direct evidence for a mechanistic explanation of *Igf2* imprinting in which
15 differential methylation of an enhancer boundary allows for epigenetic control of *Igf2* expression in the embryo (Fig. 11). In humans, a causal link between overexpression of *Igf2* and the pathogenesis of some cases of Beckwith-Wiedemann syndrome (BWS) has been suggested (Eggenchwiler, 1997; Sun, 1997; Weksberg, 1993; Joyce, 1997). BWS, or fetal overgrowth syndrome, is a disorder of prenatal
20 overgrowth and predisposition to embryonal malignancies such as Wilms tumor. Studies have shown a correlation between loss of imprinting of *Igf2* in Wilms tumor and BWS and increased methylation of the maternal *H19* allele (Steenman, 1994; Okamoto, 1997; Reik, 1995; Taniguchi, 1995). In Wilms tumor, this aberrant methylation pattern was recently shown to include the CTCF sites illustrated in
25 Figure 9a (Frevel, 1999b). These sites were consistently methylated on both alleles in Wilms tumors with loss of *Igf2* imprinting. The results described herein are consistent with the notion that the loss of *Igf2* imprinting observed in those tumors is caused by inactivation of a CTCF dependent insulator in that locus.

Recent evidence shows that in *Drosophila*, the activity of an
30 insulator can be modulated by adjacent *cis*-acting sequences (Zhou, 1999). The results described herein reveal that in vertebrates the activity of enhancer boundaries can be controlled by DNA methylation. Not relegated simply to the role of a fixed

– 41 –

boundary, some insulators may act as switches that provide a novel kind of modulated gene regulation.

References

1. Ali et al., 1994, "The Use of DNA Viruses as Vectors for Gene Therapy", *Gene Therapy*, 1:367-384.
2. G. Ascadi et al., 1994, *Human Mol. Genetics*, 3(4):579-584.
- 5 3. Baniahmad, A. et al. (1990). "Modular structure of a chicken Isozyme silencer: involvement of an unusual thyroid hormone receptor binding site". *Cell* 61, 505-514.
4. Bartolomei, M. S., Zemel, S. & Tilghman, S. M. Parental imprinting of the mouse H19 gene. *Nature* 351, 153-155 (1991).
- 10 5. Bell, A.C. and Felsenfeld, G. (1999) "Stopped at the border: boundaries and insulators". *Current Opinion in Genetics and Development*. 9, 191-198.
6. Bird (1999) *Science* 286:2287-8.
7. Bresnick, E.H. and Felsenfeld, G. (1994). "Dual promoter activation by the human beta-globin locus control region". *Proc. Natl. Acad. Sci. U.S.A.* 91, 1314-1317.
- 15 8. Bulger, M. et al. (1999). "Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes" *Proc. Natl. Acad. Sci. U.S.A.* 96, 5129-5134.
9. Burcin, M. et al., (1997). "Negative protein 1, which is required for function of the chicken Isozyme gene silencer in conjunction with hormone receptors, is identical to the multivalent zinc finger repressor CTCF". *Mol. Cell Biol.* 17, 1281-1288.
- 20 10. Chung, J.H. et al., (1993). "A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*". *Cell*. 74, 505-514.
- 25 11. Chung, J.H. et al., (1997). Characterization of the chicken beta-globin insulator". *Proc. Natl. Acad. Sci. USA*. 94, 575-580.
12. Cone, R. and Mulligan, R., 1984, Proc. Natl. Acad. Sci. USA, 81:6349.
13. DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849-859 (1991).
- 30 14. Dorsett, D. (1993). "Distance-independent inactivation of an enhancer by the suppressor of Hairy-wing DNA-binding protein of *Drosophila*". *Genetics*. 134, 1135-1144.
15. Dunaway, M. and Droge, P. (1989). "Transactivation of the *Xenopus* rRNA gene promoter by its enhancer". *Nature*. 341, 657-659.
- 35 16. Eggenschwiler, J. et al. Mouse mutant embryos overexpressing IGF-II exhibit phenotypic features of the Beckwith-Wiedemann and Simpson-Golabi-Behmel syndromes. *Genes Dev* 11, 3128-3142 (1997).

17. Evans, T. et al., (1988). "An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes". *Proc. Natl. Acad. Sci. U SA.* 85, 5976-5980.
- 5 18. Filippova, G.N. et al., (1996). "An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes". *Mol. Cell Biol.* 16, 2802-2813.
- 10 19. Frevel, M. A., Hornberg, J. J. & Reeve, A. E. A potential imprint control element: identification of a conserved 42 bp sequence upstream of H19. *Trends Genet* 15, 216-218 (1999).
20. Frevel, M. A., Sowerby, S. J., Petersen, G. B. & Reeve, A. E. Methylation sequencing analysis refines the region of H19 epimutation in Wilms tumor. *J Biol Chem* 274, 29331-29340 (1999).
- 15 21. Gdula, D.A. and Corces, V.G. (1997). "Characterization of functional domains of the Su(Hw) protein that mediate the silencing effect of mod(mdg4) mutations". *Genetics.* 145, 153-161.
- 20 22. Gdula, D.A. et al., (1996). "Genetic and molecular analysis of the gypsy chromatin insulator of *Drosophila*". *Proc. Natl. Acad. Sci. USA.* 93, 9378-9383.
23. Georgiev, P. and Kozycina, M. (1996). "Interaction between mutations in the suppressor of Hair wing and modifier of mdg4 genes of *Drosophila melanogaster* affecting the phenotype of gypsy-induced mutations". *Genetics.* 142, 425-436.
24. Gerasimova, T.I. and Corces, V.G. (1998). "Polycomb and trithorax group proteins mediate the function of a chromatin insulator". *Cell.* 92, 511 -521.
- 25 25. Gerasimova, T.I. et al., (1995). "A *Drosophila* protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation". *Cell.* 82, 587-597.
26. Geyer, P.K. and Corces, V.G. (1992). "DNA position-specific repression of transcription by a *Drosophila* zinc finger protein". *Genes Dev.* 6, 1865-1873.
- 30 27. Hark, A. T. & Tilghman, S. M. Chromatin conformation of the H19 epigenetic mark. *Hum Mol Genet* 7, 1979-1985 (1998).
28. Hebbes, T.R. et al., (1994). "Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain". *EMBO J.* 13, 1823-1830.
- 35 29. Holdridge, C. and Dorsett, D. (1991). "Repression of hsp70 heat shock gene transcription by the suppressor of hairy-wing protein of *Drosophila melanogaster*". *Mol. Cell Biol.* 11, 1894-1900.
- 40 30. Jinno, Y. et al. Mouse/human sequence divergence in a region with a paternal-specific methylation imprint at the human H19 locus. *Hum Mol Genet* 5, 1155-1161 (1996).

– 44 –

31. Kellum, R. and Elgin, S.C. (1998). "Chromatin boundaries: punctuating the genome". *Curr. Biol.* 8, R521 -4.
32. Kellum, R. and Schedl, P. (1991). "A position-effect assay for boundaries of higher order chromosomal domains". *Cell* 64, 941-950.
- 5 33. Kellum, R. and Schedl, P. (1992). "A group of scs elements function as domain boundaries in an enhancer- blocking assay". *Mol. Cell Biol.* 12, 2424-2431.
34. Khosla, S., Aitchison, A., Gregory, R., Allen, N. D. & Feil, R. Parental allele-specific chromatin configuration in a boundary- imprinting-control element upstream of the mouse H19 gene. *Mol Cell Biol* 19, 2556-2566 (1999).
- 10 35. Klenova, E.M., et al., (1993). "CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms". *Mol. Cell Biol.* 13, 7612-7624.
- 15 36. Klenova, E.M., et al., (1997). "Molecular weight abnormalities of the CTCF transcription factor: CTCF migrates aberrantly in SDS-PAGE and the size of the expressed protein is affected by the UTRs and sequences within the coding region of the CTCF gene". *Nucleic Acids Res.* 25, 466-474.
37. S. Kochanek et al., 1996, *Proc. Natl. Acad. Sci. USA*, 93:5731-5736.
- 20 38. Krebs, J.E. and Dunaway, M. (1998). "The scs and scs' insulator elements impart a cis requirement on enhancer-promoter interactions". *Mol. Cell.* 1. 301-308.
39. Labhart, P. and Reeder, R.H. (1987). "DNA sequences for typical ribosomal gene spacers from *Xenopus laevis* and *Xenopus borealis*". *Nucleic. Acids. Res.* 15, 3623-3624.
- 25 40. Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S. M. Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* 375, 34-39 (1995).
41. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting [see comments]. *Nature* 366, 362-365 (1993).
- 30 42. Lobanekov, V.V., et al. (1990). "A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'- flanking sequence of the chicken c-myc gene". *Oncogene.* 5, 1743-1753.
43. Okamoto, K., Morison, I. M., Taniguchi, T. & Reeve, A. E. Epigenetic changes at the insulin-like growth factor II/H19 locus in developing kidney is an early event in Wilms tumorigenesis. *Proc Natl Acad Sci U S A* 94, 5367-5371 (1997).
- 35 44. Pikaart, M.J., Recillas-Targa, F., and Felsenfeld, G. (1998). "Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators". *Genes Dev.* 12, 2852-2862.
- 40

- 45 -

45. Reik, W. *et al.* Imprinting mutations in the Beckwith-Wiedemann syndrome suggested by altered imprinting pattern in the IGF2-H19 domain. *Hum Mol Genet* 4, 2379-2385 (1995).
- 5 46. Reitman, M. and Felsenfeld, G. (1990). "Developmental regulation of topoisomerase II sites and DNase I-hypersensitive sites in the chicken beta-globin locus". *Mol. Cell Biol.* 10, 2774-2786.
47. Robinett, C.C., et al., (1997). "The repeat organizer, a specialized insulator element within the intergenic spacer of the *Xenopus* rRNA genes". *Mol. Cell Biol.* 17, 2866-2875.
- 10 48. Sauer, R.T., et al., (1988). "Flexibility of the yeast alpha 2 repressor enables it to occupy the ends of its operator, leaving the center free". *Genes Dev.* 2, 807-816.
49. Schmidt, J. V., Levorse, J. M. & Tilghman, S. M. Enhancer competition between H19 and Igf2 does not mediate their imprinting. *Proc Natl Acad Sci U S A* 96, 9733-9738 (1999).
- 15 50. Sinha, N.D. et al., 1984, *Nucl. Acids Res.*, 12:4539-4557
51. Spana, C., Harrison, D.A., and Corces, V.G. (1988). "The *Drosophila melanogaster* suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon". *Genes Dev.* 2, 1414-1423.
52. Stadnick, M. P. *et al.* Role of a 461-bp G-rich repetitive element in H19 transgene imprinting. *Dev Genes Evol* 209, 239-248 (1999).
- 20 53. Steenman, M. J. *et al.* Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nat Genet* 7, 433-439 (1994).
54. Sun, F. L., Dean, W. L., Kelsey, G., Allen, N. D. & Reik, W. Transactivation of Igf2 in a mouse model of Beckwith-Wiedemann syndrome [see comments]. *Nature* 389, 809-815 (1997).
- 25 55. Szabo, P. E., Pfeifer, G. P. & Mann, J. R. Characterization of novel parent-specific epigenetic modifications upstream of the imprinted mouse H19 gene. *Mol Cell Biol* 18, 6767-6776 (1998).
- 30 56. Taniguchi, T., Sullivan, M. J., Ogawa, O. & Reeve, A. E. Epigenetic changes encompassing the IGF2/H19 locus associated with relaxation of IGF2 imprinting and silencing of H19 in Wilms tumor. *Proc Natl Acad Sci U S A* 92, 2159-2163 (1995).
- 35 57. Thorvaldsen, J. L., Duran, K. L. & Bartolomei, M. S. Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. *Genes Dev* 12, 3693-3702 (1998).
58. Tremblay, K. D., Duran, K. L. & Bartolomei, M. S. A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol Cell Biol* 17, 4322-4329 (1997).

– 46 –

59. Tremblay, K. D., Saam, J. R., Ingram, R. S., Tilghman, S. M. & Bartolomei, M. S. A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat Genet* 9, 407-413 (1995).
- 5 60. Tuan, D. et al., 1985, "The β -like-globin" gene domain in human erythroid cells", *Proc. Natl. Acad. Sci. USA*, 82:6384-6388.
61. Udvardy, A. (1999). "Dividing the empire: boundary chromatin elements delimit the territory of enhancers". *EMBO J.* 18, 1-8.
- 10 62. Udvardy, A., Maine, E., and Schedl, P. (1985). "The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains". *J. Mol. Biol.* 185, 341-358.
63. Vostrov, A.A. and Quitschke, W.W. (1997). "The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation". *J. Biol. Chem.* 272, 33353-33359.
- 15 64. Webber, A. L., Ingram, R. S., Levorse, J. M. & Tilghman, S. M. Location of enhancers is essential for the imprinting of H19 and Igf2 genes. *Nature* 391, 711-715 (1998).
- 20 65. Weksberg, R., Shen, D. R., Fei, Y. L., Song, Q. L. & Squire, J. Disruption of insulin-like growth factor 2 imprinting in Beckwith- Wiedemann syndrome. *Nat Genet* 5, 143-150 (1993).
66. Yoo-Warren, H., Pachnis, V., Ingram, R. S. & Tilghman, S. M. Two regulatory domains flank the mouse H19 gene. *Mol Cell Biol* 8, 4707-4715 (1988).
- 25 67. Zhao, K., Hart, C.M., and Laemmli, U.K. (1995). "Visualization of chromosomal domains with boundary element-associated factor BEAF-32". *Cell* 81, 879-889.
68. Zhong, X.P. and Krangel, M.S. (1997). "An enhancer-blocking element between alpha and delta gene segments within the human T cell receptor alpha/delta locus". *Proc. Natl. Acad. Sci. USA.* 94, 5219-5224.
- 30 69. Zhou, J. & Levine, M. A Novel cis-Regulatory Element, the PTS, Mediates an Anti-Insulator Activity in the *Drosophila* Embryo. *Cell* 99, 567-575 (1999).

All of the references cited herein above are hereby incorporated by reference.

– 47 –

WHAT IS CLAIMED IS:

1. An isolated DNA molecule comprising SEQ ID NO:1, said DNA molecule having enhancer-blocking function.
2. An expression construct comprising the DNA molecule of
5 claim 1.
3. The construct according to claim 2, wherein the construct is operable when inserted into the DNA of a cell to insulate the expression of one or more genes from one or more *cis*-acting regulatory sequences in chromatin.
4. The construct according to claim 3, wherein said cell is a
10 mammalian cell.
5. A vector construct comprising:
 - (a) the DNA molecule according to claim 1;
 - (b) a promoter domain;
 - (c) a gene operably linked to the promoter domain; and
 - 15 (d) an enhancer domain 5' of the promoter domain, wherein the insulator DNA molecule is positioned between the enhancer and the promoter domains so as to operably insulate the transcription and expression of the gene from *cis*-acting regulatory elements in chromatin.
6. An isolated DNA construct for incorporation into a host cell
20 and for insulation of the expression of a gene therein, comprising:
 - a) DNA comprising a transcription unit comprising an expressible gene, a promoter to drive the transcription of the gene, and an enhancer element; and
 - b) one or more DNA molecules according to claim 1, the
25 molecules being positioned in sufficient proximity to the transcription unit and to the gene to insulate the transcription and expression of the gene from *cis*-acting DNA regulatory sequences in chromatin outside of the DNA according to a).

– 48 –

7. The DNA construct according to claim 6, wherein the expressible gene is a structural gene.

8. The DNA construct according to claim 6, wherein the expressible gene is selected from the group consisting of protein-encoding genes, hormone-encoding genes, peptide hormone-encoding genes, enzyme-encoding genes, and antibiotic-resistance-encoding genes.

9. The DNA construct according to claim 8, wherein the expressible gene is a neomycin-resistance gene or a hygromycin-resistance gene.

10. A mammalian cell stably transfected with the construct according to claim 2.

11. A method of insulating the expression of an introduced gene from cis-acting DNA regulatory sequences in the chromatin into which the gene has integrated, comprising:

a) introducing into a cell the DNA construct according to claim 2;

b) integrating the construct into the chromatin of the cell, wherein the expression of a resultant integrated heterologous gene is insulated from cis-acting DNA regulatory sequences in the chromatin of said cell.

12. The method according to claim 11, further comprising introducing into the cell a DNA construct containing a gene encoding the CTCF protein, wherein CTCF is expressed in the cell.

13. A kit for insulating the expression of a transfected and expressed gene, comprising a vector comprising the insulator molecule according to claim 1.

14. A pharmaceutical composition comprising the construct according to claim 2 in a pharmaceutically acceptable diluent, carrier, or excipient.

15. A method of blocking activity of an enhancer of a gene in a cell, comprising:

– 49 –

- a) introducing into the cell a construct containing a DNA molecule comprising SEQ ID NO:1;
- b) introducing into the cell a construct containing a gene encoding the CTCF protein, wherein the CTCF protein is
5 expressed in the cell and binds to the DNA molecule of step (a).
16. An isolated DNA molecule comprising the sequences shown in SEQ ID NOS:84-87, said DNA molecule having enhancer-blocking function.
17. An isolated DNA molecule comprising the sequences shown in SEQ ID NOS:88-91.
- 10 18. An isolated DNA molecule comprising the sequences shown in SEQ ID NOS:92-98.
19. The DNA molecule according to claims 16, 17 or 18, wherein the molecule contains binding site for the CTCF protein.
20. The DNA molecule according to claims 16, 17 or 18, wherein
15 the enhancer-blocking activity of the molecule is dependent upon CTCF binding to the molecule.
21. The DNA molecule according to claims 16, 17 or 18, wherein methylation of the cytosines (C) of the CpG residues in the molecule prevents CTCF binding to the molecule and inhibits the enhancer-blocking function of the
20 molecule.
22. A method of activating the expression of an introduced gene from cis-acting DNA regulatory sequences in the chromatin into which the gene has integrated, comprising:
- a) introducing into a cell a first DNA construct comprising a
25 transcription unit comprising an expressible gene, a promoter to drive the expression of the gene, an enhancer element and an insulator element;
- b) introducing into a cell a second DNA construct encoding a fusion protein, said fusion protein comprising the enzymatic domain of a methylase and the DNA binding domain of a DNA binding protein.

– 50 –

23. The method according to claim 22, wherein the first DNA construct further comprises DNA binding sequences for the DNA binding protein encoded by the second construct.

5 24. The method according to claim 22, wherein the DNA binding protein is Gal4 or LexA.

25. The method according to claim 22, wherein the methylase methylates the cytosines of the CpG residues in the insulator element.

26. The method according to claim 22, wherein the methylase is DNA methyltransferase 3.

10 27. The method according to claim 22, wherein the insulator element comprises sequences selected from the group consisted of SEQ ID NOS:84-100.

SEQUENCE LISTING

<110> Bell, Adam

West, Adam

Felsenfeld, Gary

<120> DNA BINDING PROTEIN AND SEQUENCE AS INSULATORS HAVING
SPECIFIC ENHANCER BLOCKING ACTIVITY FOR REGULATION OF
GENE EXPRESSION

<130> 20264297PC

<140> TBA

<141> 2000-04-19

<150> 60/141,728

<151> 1999-06-30

<160> 100

<170> PatentIn Ver. 2.1

<210> 1

<211> 42

<212> DNA

<213> CHICKEN

<220>

<400> 1

cccagggatg taattacgtc cctccccgc tagggggcag ca

42

<210> 2

<211> 34

<212> DNA

<213> CHICKEN

<220>

<400> 2

gggatgtaat tacgtccctc ccccgctagg gggc

34

<210> 3

<211> 11

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 3

ccatacgttr y

11

<210> 4

<211> 8

<212> DNA

<213> Artificial Sequence

<220>

<220>

<223> Description of Artificial Sequence:primer

<400> 4

gggcgggg

8

<210> 5

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<220>

<223> Description of Artificial Sequence:primer

<400> 5

tacattaatg ca

12

<210> 6

<211> 42

<212> DNA

<213> CHICKEN

<400> 6

cccatttcgt gccggccgtc cctccccgc tagggggcag ca

42

<210> 7

<211> 42

<212> DNA

<213> CHICKEN

<400> 7

cccagggatg taattaatga aagaaccgc tagggggcag ca

42

<210> 8

<211> 42

<212> DNA

<213> CHICKEN

<400> 8

cccagggatg taattacgtc cctccaaata gctttttcag ca

42

<210> 9

<211> 42

<212> DNA

<213> CHICKEN

<400> 9

acgacggggg atcgccccct ccctgcatta atgtaggac cc

42

<210> 10

<211> 23

<212> DNA

<213> CHICKEN

<400> 10

gtaattacgt ccctcccccg cta

23

<210> 11

<211> 42

<212> DNA

<213> CHICKEN

<400> 11

cccaggtcgg taattacgtc cctccccgc tagggggcag ca

42

<210> 12

<211> 42

<212> DNA

<213> CHICKEN

<400> 12

cccagggatt gcattacgtc cctccccgc tagggggcag ca

42

<210> 13

<211> 42

<212> DNA

<213> CHICKEN

<400> 13

cccagggatg tacggacgtc cctccccgc tagggggcag ca

42

<210> 14

<211> 42

<212> DNA

<213> CHICKEN

<400> 14

cccagggatg taattacgtc cctccaaagc tagggggcag ca

42

<210> 15

<211> 42

<212> DNA

<213> CHICKEN

<400> 15

cccagggatg taattacgtc cctcccccta gagggggcag ca

42

<210> 16

<211> 42

<212> DNA

<213> CHICKEN

<400> 16

cccagggatg taattacgtc cctccccgc tcttgggcag ca

42

<210> 17

<211> 42

<212> DNA

<213> CHICKEN

<400> 17

cccagggatg taattacgtc cctccccgc taggtttcag ca

42

<210> 18

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 18

cgcggggctcc gtgagcgggg agggcgcgcc gcgagggggc ggcc

44

<210> 19

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 19

agcgggcgca gttccccggc ggcgccgcta ggggtctctc

40

<210> 20

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 20

caaaaagaca tgtaaatacc atagctatcc agtagaggtc tcaa

44

<210> 21

<211> 47

<212> DNA

<213> Xenopus laevis

<400> 21

accgattcg gggtcggggc cccgggggtg cccgctaagg ggccccg

47

<210> 22

<211> 45

<212> DNA

<213> *Xenopus laevis*

<400> 22

cccgattcgg gggtcggggc cccgggggtg ccgcggggg cccgg

45

<210> 23

<211> 45

<212> DNA

<213> *Xenopus laevis*

<400> 23

accgattcgg gggtcggggc cccgggggtg ccgcggggg ccccg

45

<210> 24

<211> 45

<212> DNA

<213> *Xenopus laevis*

<400> 24

accgattcgg gggtcggggc cccgggggtg ccgcggggg ccccg

45

<210> 25

<211> 42

<212> DNA

<213> *Xenopus laevis*

<400> 25

accgattcgg gggtcggggc cccggggccc gcggggggccc cg

42

<210> 26

<211> 45

<212> DNA

<213> *Xenopus laevis*

<400> 26

accgattcgg gggtcggggc cccgggggtg ccgcggggg ccccg

45

<210> 27

<211> 47

<212> DNA

<213> *Xenopus laevis*

<400> 27

acccgattcg gggcggggc cccgggggtg cccgctaagg ggccccg

47

<210> 28

<211> 39

<212> DNA

<213> Homo sapiens

<400> 28

cccaggcctg cactgccgcc tgccggcagg ggtccagtc

39

<210> 29

<211> 33

<212> DNA

<213> CHICKEN

<400> 29

aggcggcct gggagctcac ggggacagcc ccc

33

<210> 30

<211> 33

<212> DNA

<213> CHICKEN

<400> 30

aggcggcct gggagcgccg gaccggagcg gag

33

<210> 31

<211> 33

<212> DNA

<213> CHICKEN

<400> 31

aggcggcct gctccgctcc ggtccggcgc tcc

33

<210> 32

<211> 34

<212> DNA

<213> CHICKEN

<400> 32

aggcggcct gtcattctaa atctctcttt cagc

34

<210> 33
<211> 33
<212> DNA
<213> CHICKEN

<400> 33
aggcgcgccg cccccagga tgtaattacg tcc

33

<210> 34
<211> 45
<212> DNA
<213> CHICKEN

<400> 34
agcccccccc caaagcccc agggatgggg gcagcagcga gccgc

45

<210> 35
<211> 45
<212> DNA
<213> CHICKEN

<400> 35
ggcggctcgc tgctgcccc atccctgggg gctttggggg ggggc

45

<210> 36
<211> 25
<212> DNA
<213> CHICKEN

<400> 36
ccgagccggc agcgtgcggg gacag

25

<210> 37
<211> 40
<212> DNA
<213> CHICKEN

<400> 37
ccgcacgct gccggctcgg cggaccggag cggagccccg

40

<210> 38
<211> 25
<212> DNA

<213> CHICKEN

<400> 38

cctctgaacg cttctcgctg ctctt

25

<210> 39

<211> 40

<212> DNA

<213> CHICKEN

<400> 39

cagcgagaag cgttcagagg cttccccgt gcccgggctg

40

<210> 40

<211> 36

<212> DNA

<213> CHICKEN

<400> 40

aggcgcgccg cccaggtgtc tgcaggctca aagagc

36

<210> 41

<211> 39

<212> DNA

<213> CHICKEN

<400> 41

aggcgcgccg aattccagaa atctttgatt tcagatgct

39

<210> 42

<211> 40

<212> DNA

<213> CHICKEN

<400> 42

aggcgcgccg gatccactc ttagccatta tactgcattg

40

<210> 43

<211> 48

<212> DNA

<213> CHICKEN

<400> 43

tgagcatctt cagggcccct ggattccatt tcagagcttc cggttctc

48

<210> 44

<211> 24

<212> DNA

<213> CHICKEN

<400> 44

atccaggggc cctgaagatg ctca

24

<210> 45

<211> 107

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 45

aggcgcgccg ggatgtaatt acgtccctcc cccgctaggg ggcagcagcg agcgcccggg 60
gctccgctcc ggtccggcgc tcccccgca tccccgaggg cgcgcct 107

<210> 46

<211> 108

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 46

aggcgcgccg ggatgtaatt acgtccctaa cccgctaggg ggcagcagcg agccgaacgg 60
ggctccgctc cgggccggcg ctaaccccg ctaaccccgagg cgcgcct 108

<210> 47

<211> 104

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 47

aggcgcgccg ggatgtacgt cctcccccg ctagggggca gcagcgagcc gcccggggct 60

ccgctccggt ccggcgctcc ccccgcatcc ccgagggcgc gcct

104

<210> 48

<211> 94

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 48

aggcgcgccc ccagggatgt aattacgtcc ctcccccgct agggggcagc accggtccgg 60
cgctcccccc gcatccccga gccggggcgc gcct 94

<210> 49

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 49

aggcgcgccg ggggcagcag cgagccgccc ggggctccgc tccggtccgg cgctcccccc 60
gcatccccga gggcgcgctt 80

<210> 50

<211> 89

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 50

aggcgcgccc caaagccccc agggatgtaa ttacgtccct ccccgctag ggggcagcag 60
cgagccgccc ggggctccgc ggcgcgcct 89

<210> 51

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 51

aggcgcgccc ccagggatgt aattacgtcc ctcccccgct agggggcagc agcgcgcct 60

<210> 52

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 52

aggcgcgccc cggcccgcg ctccccccgc atccccgagc cggggcgcg ct 52

<210> 53

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 53

aggcgcgcca aaatacattg cataccctct ttaataaaa aatattgcat acgttgacga 60
aacaatttt cgttgcatac ccaataaaaag gcgcgcct 98

<210> 54

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 54

aggcgcgccg ggggggggca cggagcccct cggccgcccc ctgcgggcgc gccctccccg 60
ctcacggagc ccgcgcggag ccgggggcga ggcgcgcc 98

<210> 55

<211> 97

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 55

aggcgcgcct ttagctgcat ttgacatgaa gaaattgaga cctctactgg atagctatgg 60
tatttacatg tctttttgct tagttactag gcgcgcc 97

<210> 56

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 56

aggcgcgccc cctccccggc cgagcgggag cagttccccg gcggcgccgc taggggtctc 60
tctcgggtgc cgagcggggt gggccggata ggcgcgcc 98

<210> 57

<211> 99

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 57

aggcgcgcgc gggacccgat tcggggtcgg ggccccgggg gtgcccgccta aggggccccg 60
gggggccctc ccggcgaaga ggggccatt ggcgcgcct 99

<210> 58

<211> 137

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 58

ggcgcgcgcgt ggaagaggga tggtgagggc ccaggggctg ccttgccggt gcattggctg 60
cccaggcctg cactgccgcc tgccggcagg ggtccagtc acgagacca gtcacctgct 120
ggcggaaggg cgcgcct 137

<210> 59

<211> 16

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 59

Ala Pro Asn Gly Asp Leu Thr Pro Glu Met Ile Leu Ser Met Met Asp

1

5

10

15

<210> 60

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 60

cccagggatg taattacgtc cctcccccg c tagggggcag caggcgcgcc t

51

<210> 61

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 61

aggcgcgcc t gctgc

15

<210> 62

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 62

aggcgcgcca agctttgtca cagcggaacc caacctatg

39

<210> 63

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 63

aggcgcgccc agagctcttt ctccaccact tgtctaagt 39

<210> 64

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 64

aggcgcgccg gtacctcgtg gactcggact cccaaatca 39

<210> 65

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 65

aggcgcgcca tagtagctat acttcaattt tca 33

<210> 66

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 66

aggcgcgctt ttataagagg ttggaacact tgt 33

<210> 67

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 67

ccctattctt ggacgtctgc tgaatctatt ggaattcaca aatggcaatg c 51

<210> 68

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 68

gattcagcag acgtccaaga ataggg 26

<210> 69

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 69

gactcggact cccaaatcaa caaggacgga ttgcaactga ttgagttttc 50

<210> 70

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 70

ccttgttgat ttgggagtc gagtc 25

<210> 71

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 71

aggcgcgccca agactgaagg agctacccaa gaa

33

<210> 72

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 72

aggcgcgcct ttataagagg ttggaacact tgt

33

<210> 73

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 73

aggcgcgccca gagaacttga ctcattccct acac

34

<210> 74

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 74

agaagctgtt atgtgcaaca agggagcgat tcattcccag caatatcc

48

<210> 75
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 75
cccttggtgc acataacagc ttct 24

<210> 76
<211> 88
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 76
aggcgcgccg ttgtgggggt tatacgcggg agttgccgcg tggtaggcagc aaaatcgatt 60
gcgccaaacc taaagagccg gcgcgcct 88

<210> 77
<211> 83
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 77
aggcgcgcc aatcctttgtg tgtaaagacc agggttgccg cacggcggca gtgaagtctc 60
gtacatcgca gtccggcgcg cct 83

<210> 78
<211> 88
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Primer

<400> 78

aggcgcgccc tggtatgtgc aacaagggaa cggatgctac cgcgcggtgg cagcatactc 60
ctatatatcg tggcccaaag gcgcgcct 88

<210> 79

<211> 88

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 79

aggcgcgcca cgctgtgcag atttggctat agctaaatgg acagacgatg ccgcgtggtg 60
gcagtacaat actacatatg gcgcgcct 88

<210> 80

<211> 91

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 80

gccctgatgg cgcagaatcg gctgtacgtg tggaatcaga agtggccgcg cggcggcagt 60
gcaggctcac acatcacagc ccgagcacgc c 91

<210> 81

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 81

aggcgcgccc ccagggatgt aattacgtcc ctccaaatag ctttttcagc aggcgcgcct 60

<210> 82

<211> 88

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Primer

<400> 82

aggcgcgcct gctgaatcag ttgtgggggt tatacgcggg agttgaatat gttgttactc 60
aaaatcgatt gcgcaaacg gcgcgcct 88

<210> 83

<211> 102

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 83

gctgttatgt gcaacaaggg aacggatgct taccgcgcgg tggcagcata ctcctatata 60
tcgtggccca aatgctgcca acttgggggg agcgattcat tc 102

<210> 84

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 84

gttgtgggggt ttatacgcgg gagttgccgc gtggtggcag caaaatcg 48

<210> 85

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 85

tttgtgtgta aagaccaggg ttgccgcacg gcggcagtga agtct 45

<210> 86

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 86

tatgtgcaac aagggaaacgg atgctaccgc gcggtggcag catactcc 48

<210> 87

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 87

gctatagcta aatggacaga cgatgccgcg tgggtggcagt acaatac 47

<210> 88

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 88

ttgtgtggtt taaaacgcgg aagttgccgc gtggtggcag caaaaatc 48

<210> 89

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 89

tcctttgcgc gtaaaaacca ggcctgccgc gtggcggcag tgaagtcg 48

<210> 90

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 90

ttgtgtgcac ggggaaatgg atgttaccgc gcggtggcag catactcc 48

<210> 91

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 91

tgactatagc tagatggaca aatatgccgc gtggtggcag tacaacc 48

<210> 92

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 92

ggctgtacgt gtggaatcag aagtggccgc gcggcggcag tgcaggct 48

<210> 93

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 93

ggttgtagtt gtggaatcgg aagtggccgc gcggcggcag tgcaggct 48

<210> 94

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 94

ggttgtagct gtggaatcgg aagtggccgc gtggcggcag tgcaggct 48

<210> 95

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 95

ggttgtaagt gtggactcaa aagtggccgc gcggcggcag tgcaggct 48

<210> 96

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 96

ggttgtagtt gtggaatcgg aggtggctgc gcggcggcag tgcaggct 48

<210> 97

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 97

ggttgtagtt gtggaatcgg aagtggccgc gcggcggcag tgcaggct 48

<210> 98

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 98

ggttgtggct gtggagacgg aaatggccga gaggcggcag tggtgact 48

<210> 99

<211> 14

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<220>

<223> The "n" at positions 6 and 9 can be either C or T.

<400> 99

ccgcgnggng gcag 14

<210> 100

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 100

cccagggatg taattacgtc cctccccgc tagggggcag cagcgagc 48

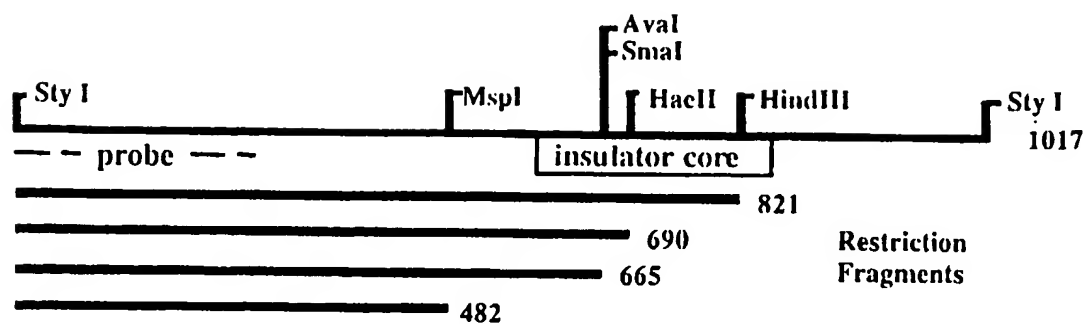


FIG. 1A

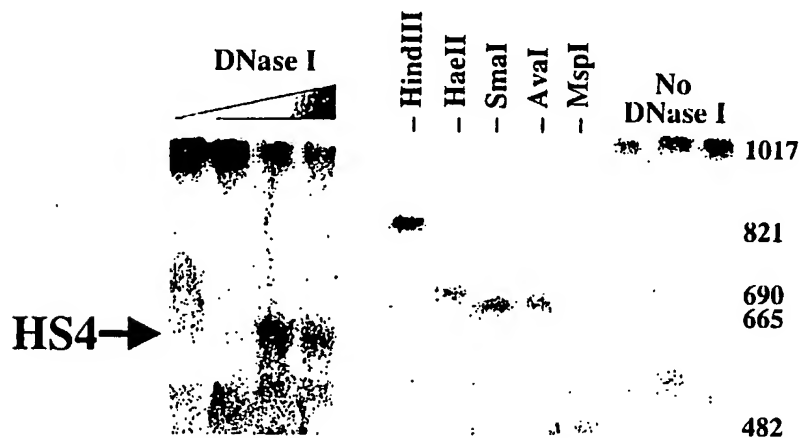


FIG. 1B

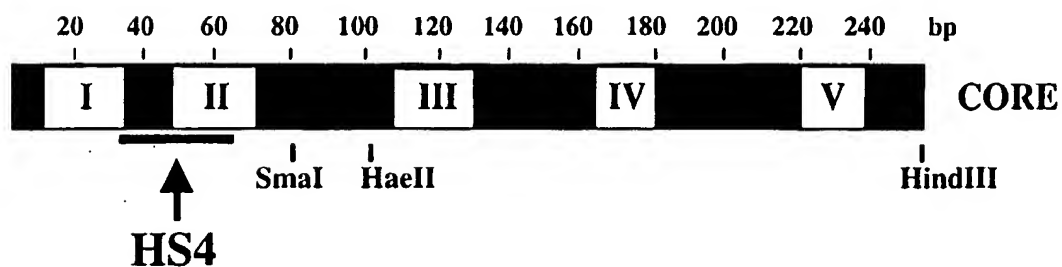
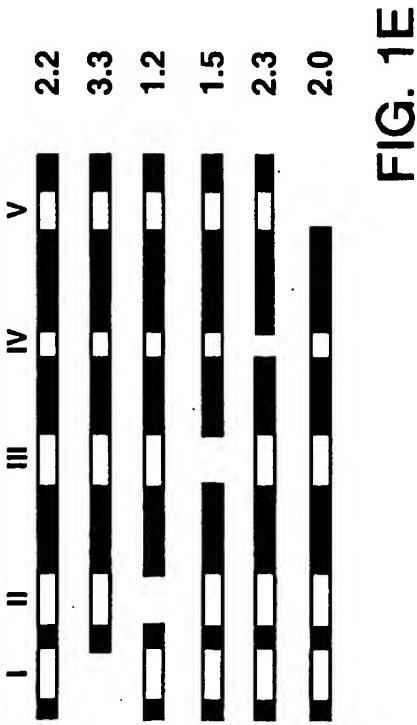
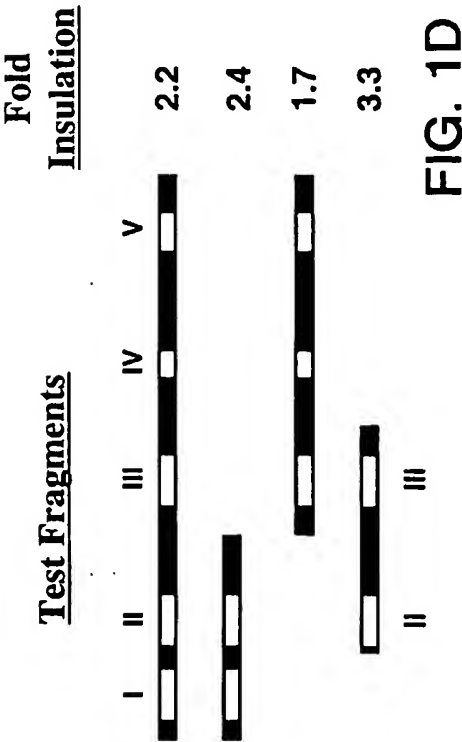
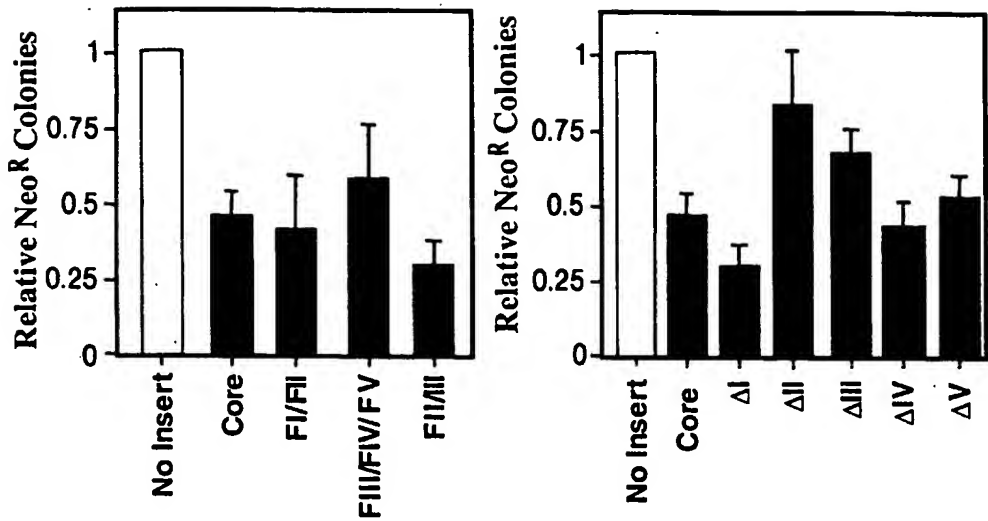
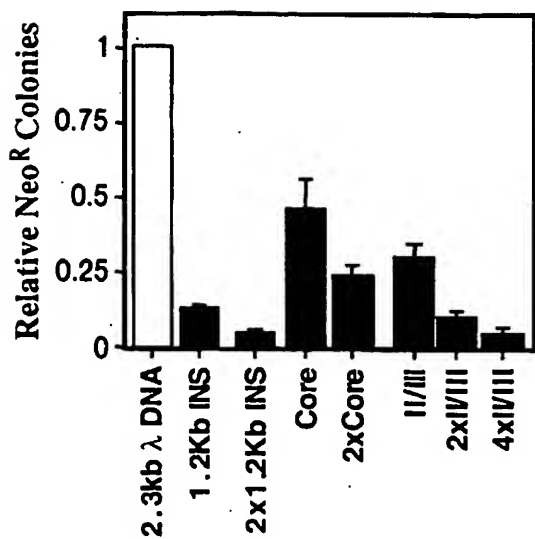


FIG. 1C





**Fold
Insulation**

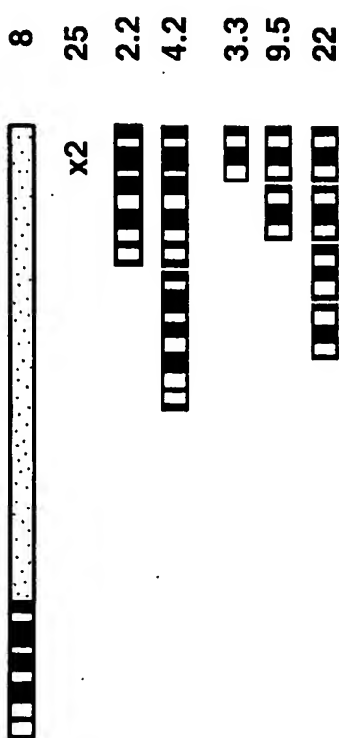


FIG. 1F

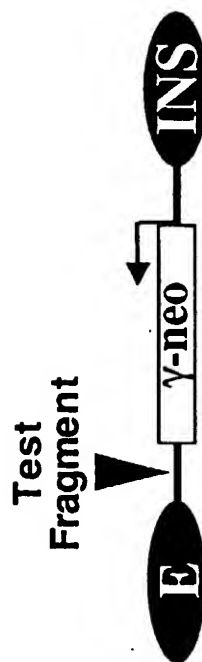
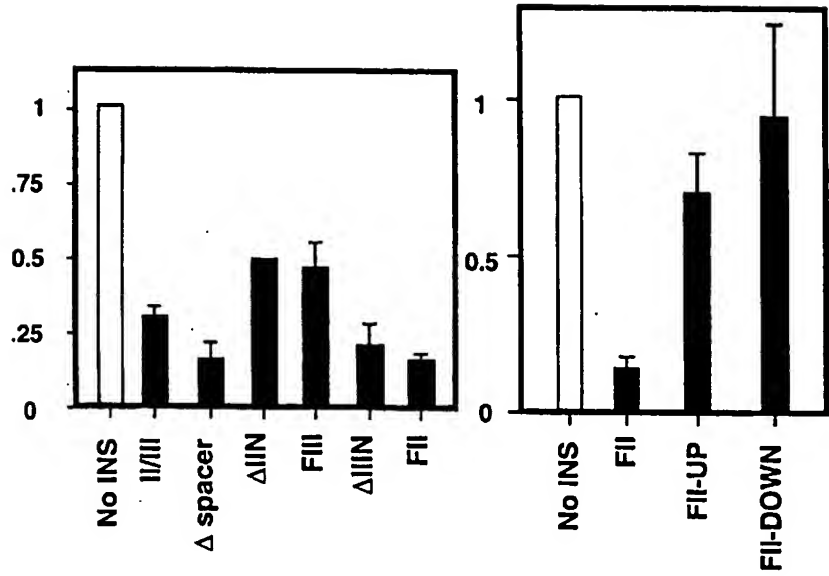


FIG. 1G



Fold Insulation

3.3
6.4
2.0
2.2
5.0
6.0

FIG. 2A

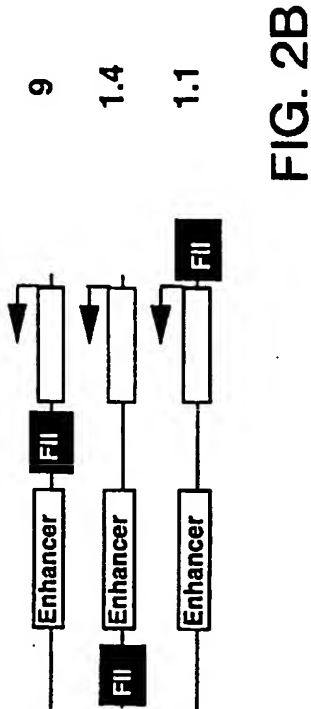
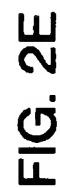
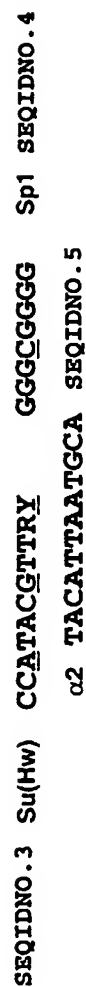
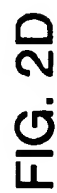
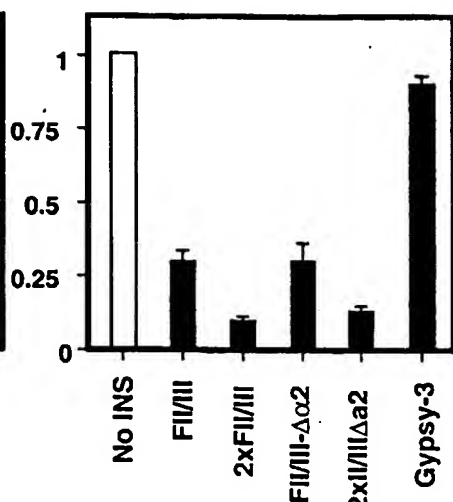
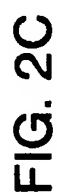
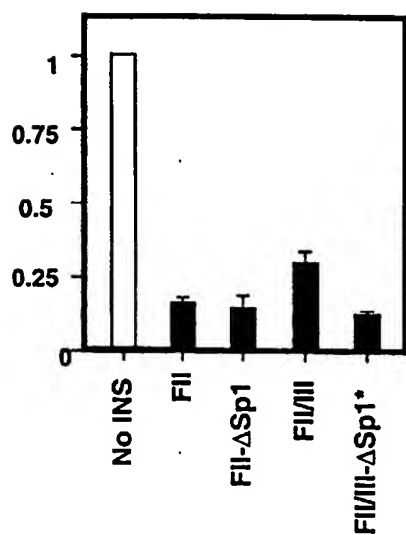
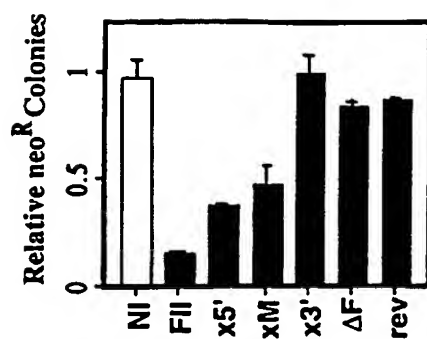


FIG. 2B



6/20

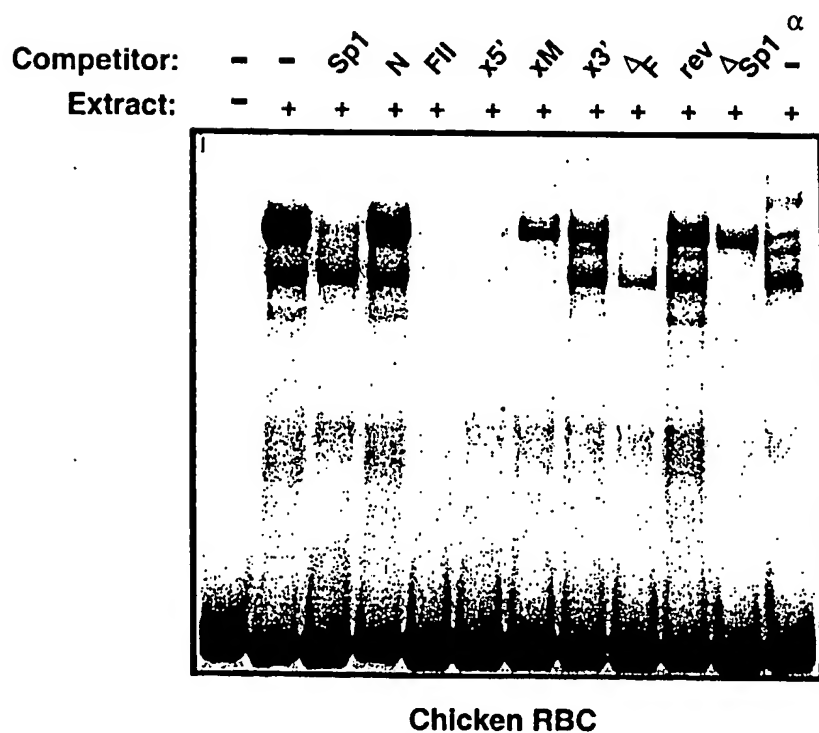
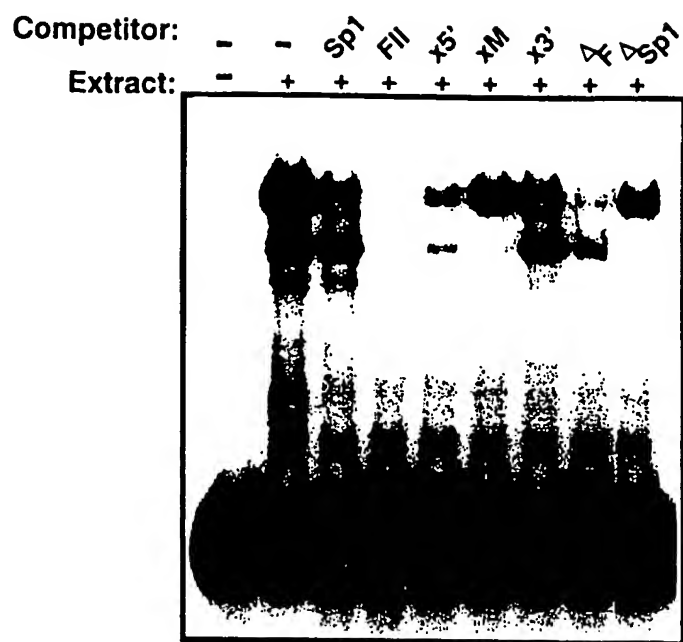


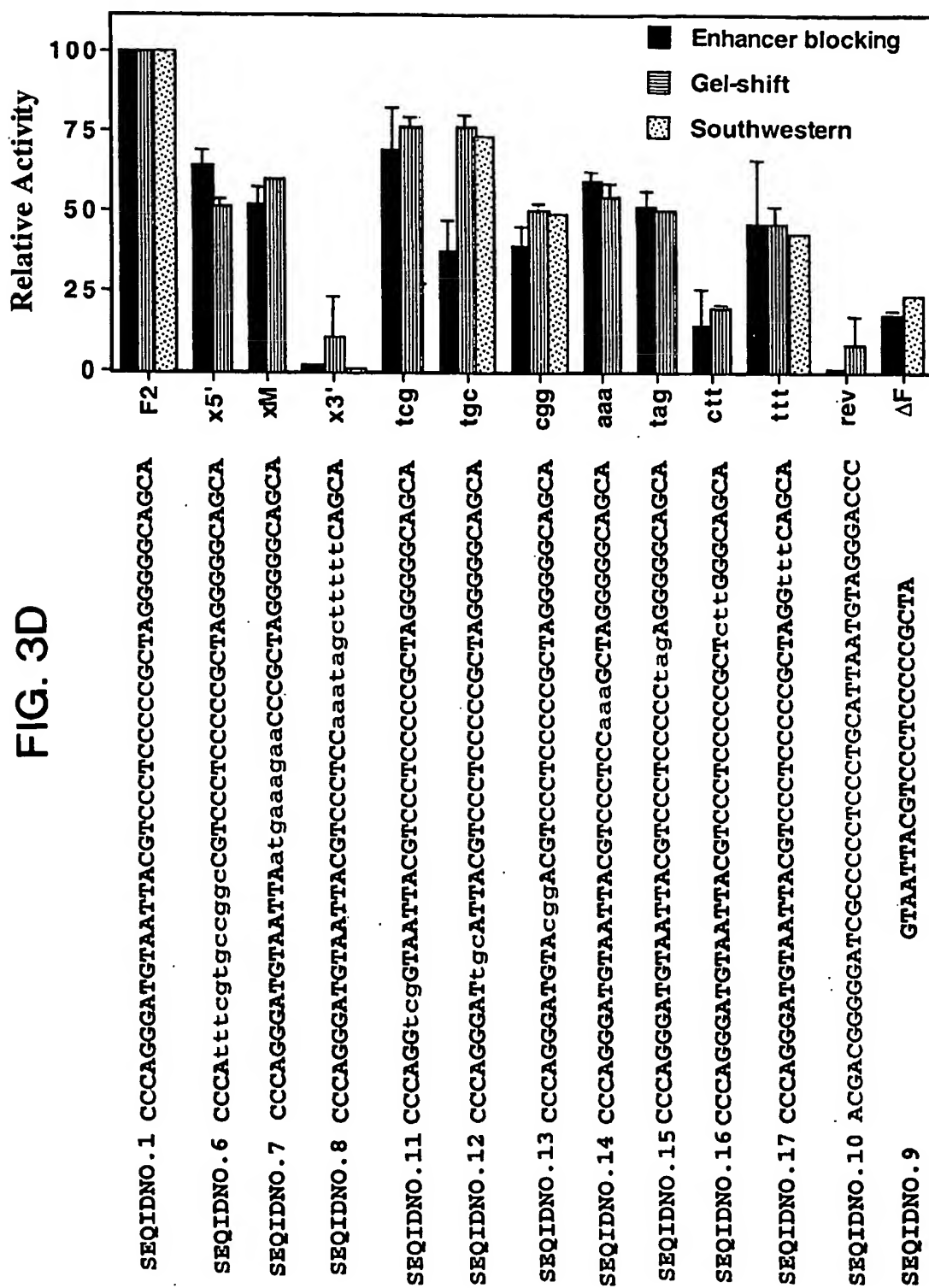
Fold
Insulation

- 6 3 2 - 1.2 1.1

SEQIDNO.1 CCCAGGGATGTAATTACGTCCCTCCCCCGCTAGGGGGCAGCA
 SEQIDNO.6 CCCAtttcgtgcggccCGTCCCTCCCCCGCTAGGGGGCAGCA
 SEQIDNO.7 CCCAGGGATGTAATTAAatgaaagaaCCCCGCTAGGGGGCAGCA
 SEQIDNO.8 CCCAGGGATGTAATTACGTCCCTCCaaatagctttttCAGCA
 SEQIDNO.9 GTAAATTACGTCCCTCCCCCGCTA
 SEQIDNO.10 ACGACGGGGGATCGCCCCCTCCCTGCATTAATGTAGGGACCC

FIG. 3A





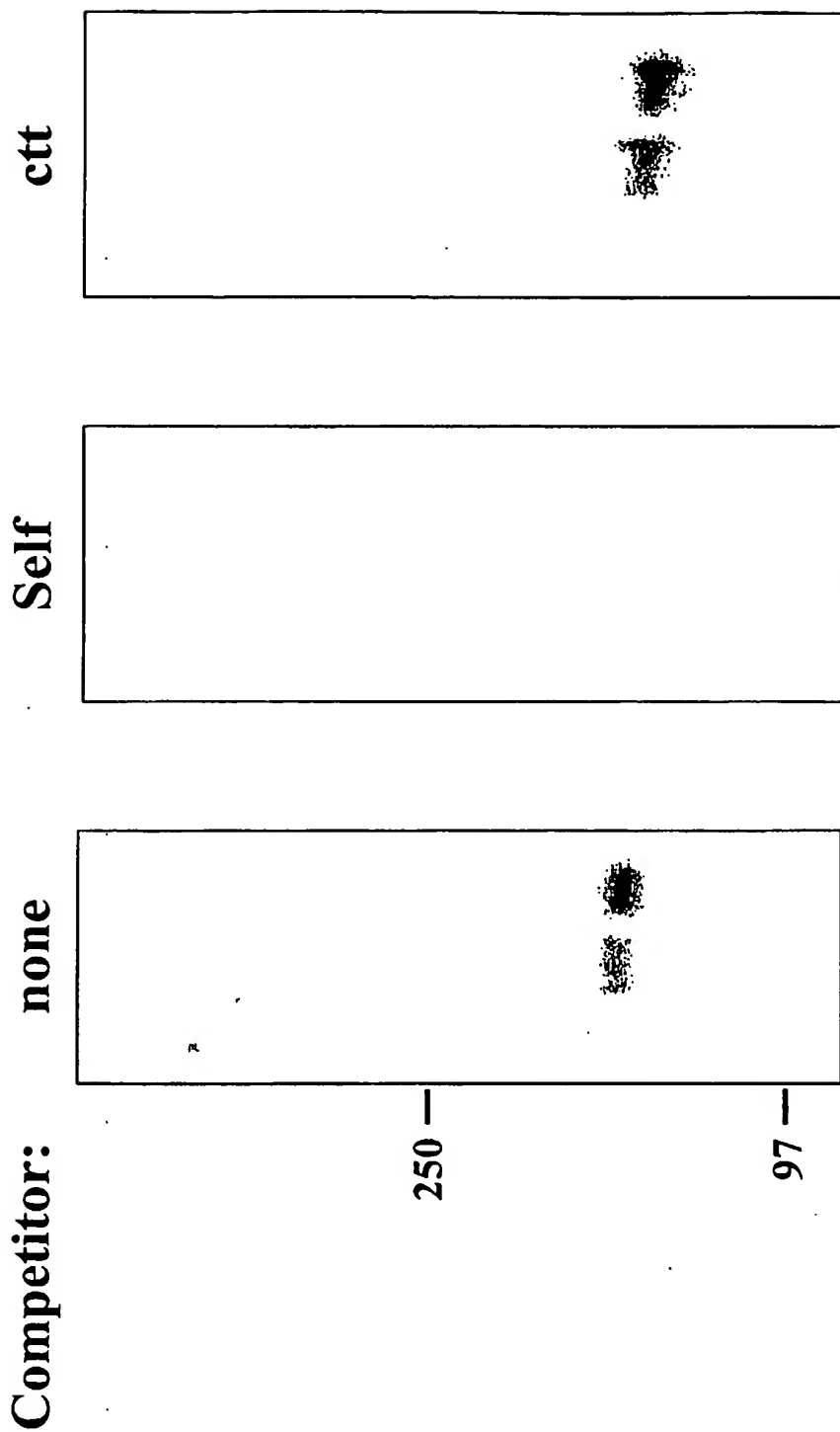


FIG. 4A

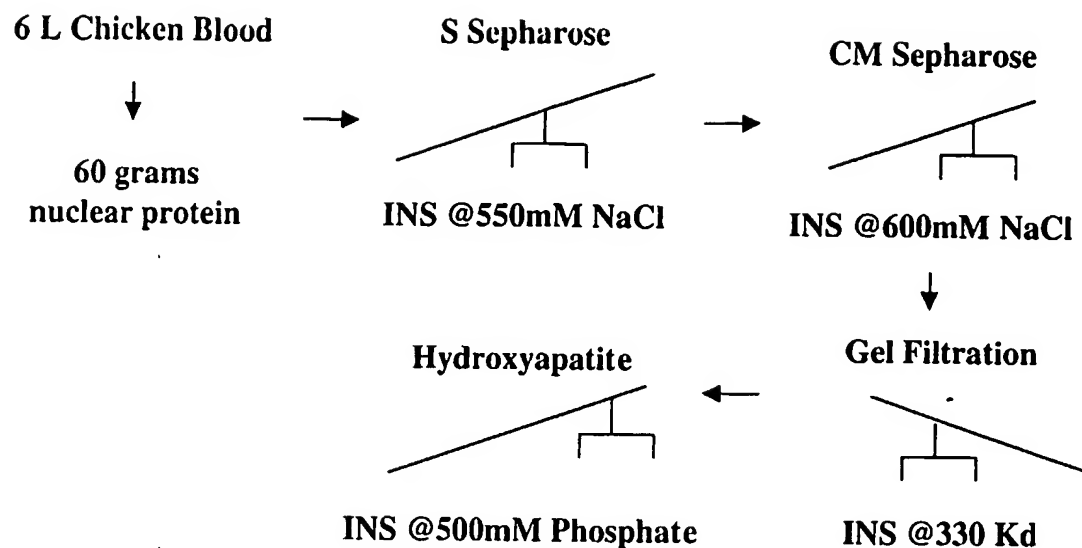
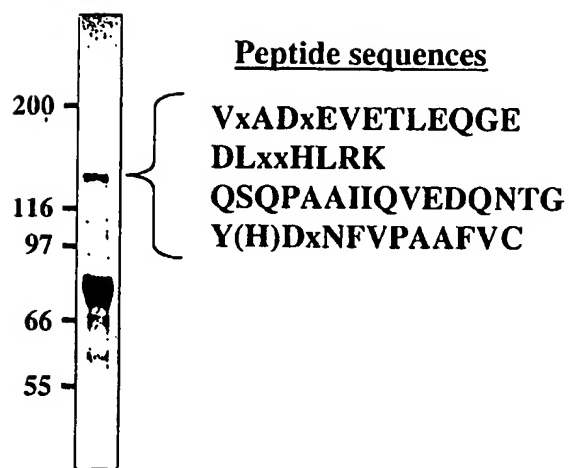
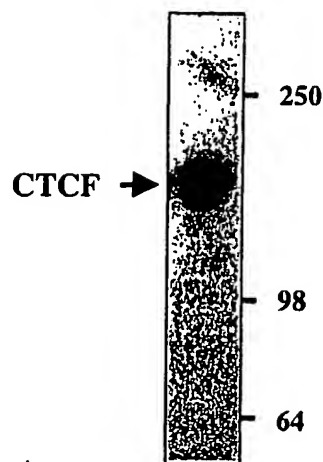


FIG. 4B



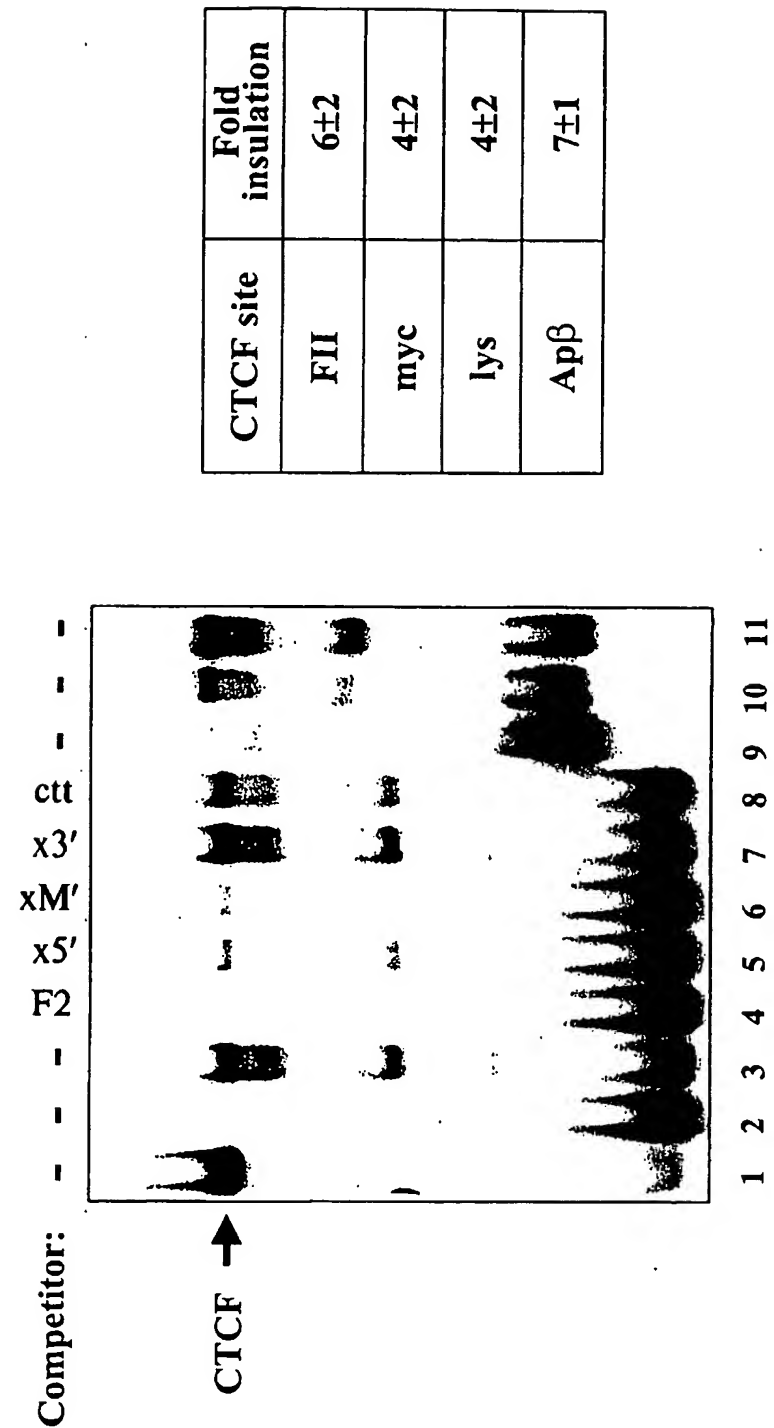
Coomassie

FIG. 4C



Southwestern

FIG. 4D



CTCF site	Fold insulation
FII	6±2
myc	4±2
lys	4±2
Apβ	7±1

FIG. 5

FII CCCAGGG--ATGTAATTACGTCCCTCCCGGCTAGGGGAGCA SEQIDNO.1
 mycFV CGCGGCTCCGTAGCGGGAGGGCGCGCCGCGAGGGGGGGCC SEQIDNO.18
 lys AGGGGCGGAGTCCCGGGCG---CGCCGCTAGGGTCTCTC SEQIDNO.19
 ap β CAAAAGACATGTAATACCATAGCTATCCAGTAGGGTCTCAA SEQIDNO.20
 FIIX3' CCCAGGG--ATGTAATTACGTCCCTCCaaatagctttttcAGCA SEQIDNO.8

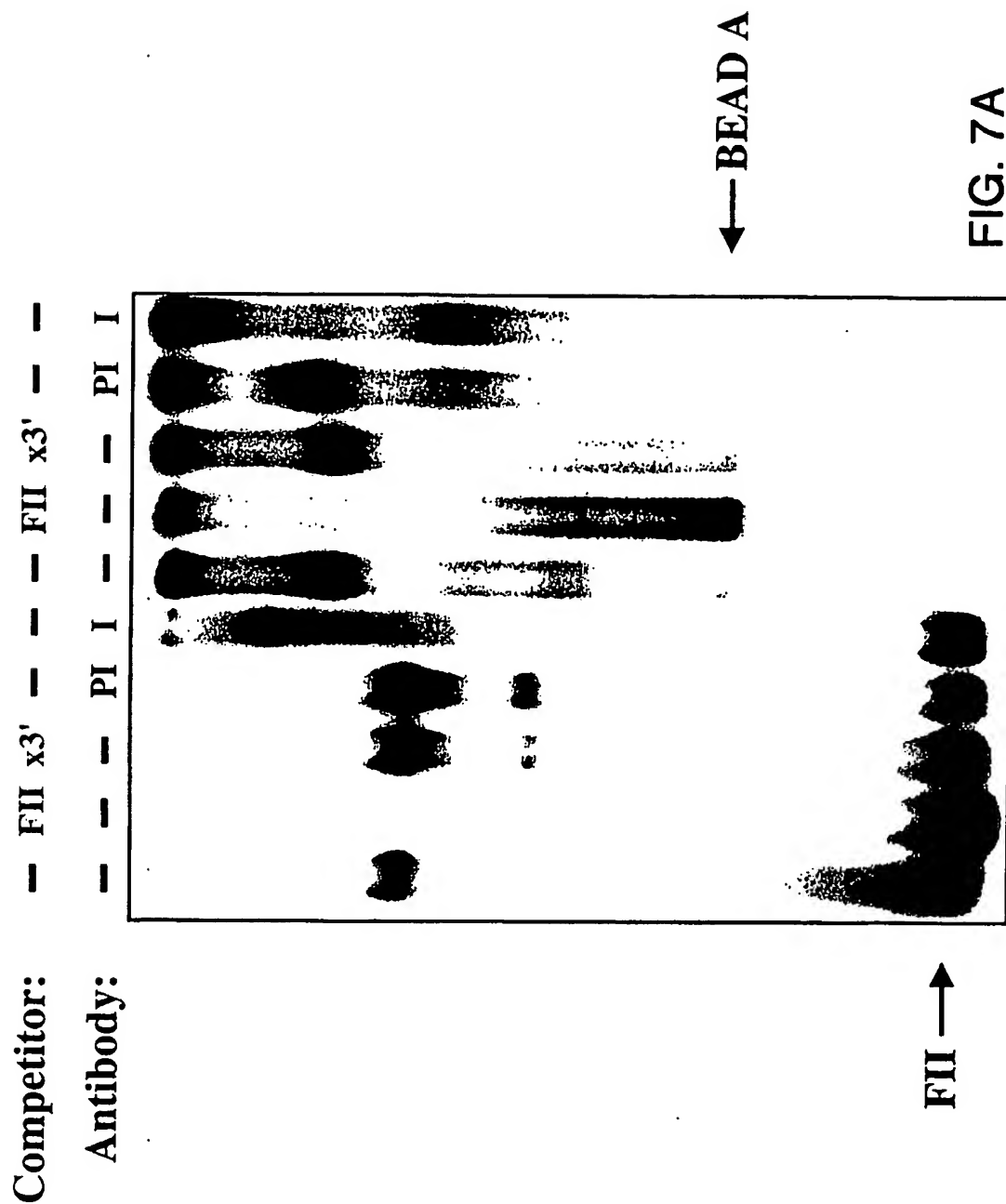
FIG. 6A

FII CCCAGGATGTAAATTACGTCCCT-----CCCCGCTAGGGGAGCA SEQIDNO.1
 RO 1...ACCGATTCCGGGTCGGGCCCCCGGGGTGCCCGCTAGGGGCCCG...SEQIDNO.21
 RO 2...CCGATTCGGGTCGGGCCCCGGGGTGCCCGC--GGGGGCCCGG...SEQIDNO.22
 RO 3...ACCGATTCCGGGTCGGGCCCCCGGGGTGCCCGC--GGGGGCCCGG...SEQIDNO.23
 RO 4...ACCGATTCCGGGTCGGGCCCCCGGGGTGCCCGC--GGGGGCCCGG...SEQIDNO.24
 RO 5...ACCGATTCCGGGTCGGGCCCCCGGG--CCCCGC--GGGGGCCCGG...SEQIDNO.25
 RO 6...ACCGATTCCGGGTCGGGCCCCCGGGGTGCCCGC--GGGGGCCCGG...SEQIDNO.26
 RO 7...ACCGATTCCGGGTCGGGCCCCCGGGGTGCCCGCTAGGGGCCCGG...SEQIDNO.27

FIG. 6B

FII CCCAGGATGTAAATTACGTCCCTCCCGGCTAGGGGAGCA SEQIDNO.1
 BEAD A CCCAGGCCTGCACTGCCG--CCTGCCG-GCAGGGTCCAGTC SEQIDNO.28

FIG. 6C



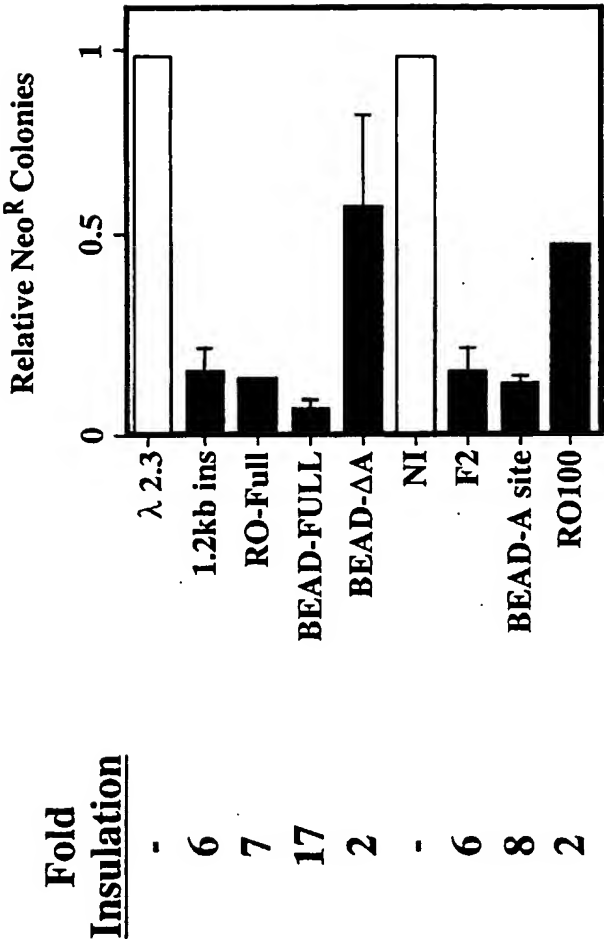


FIG. 7B

15/20

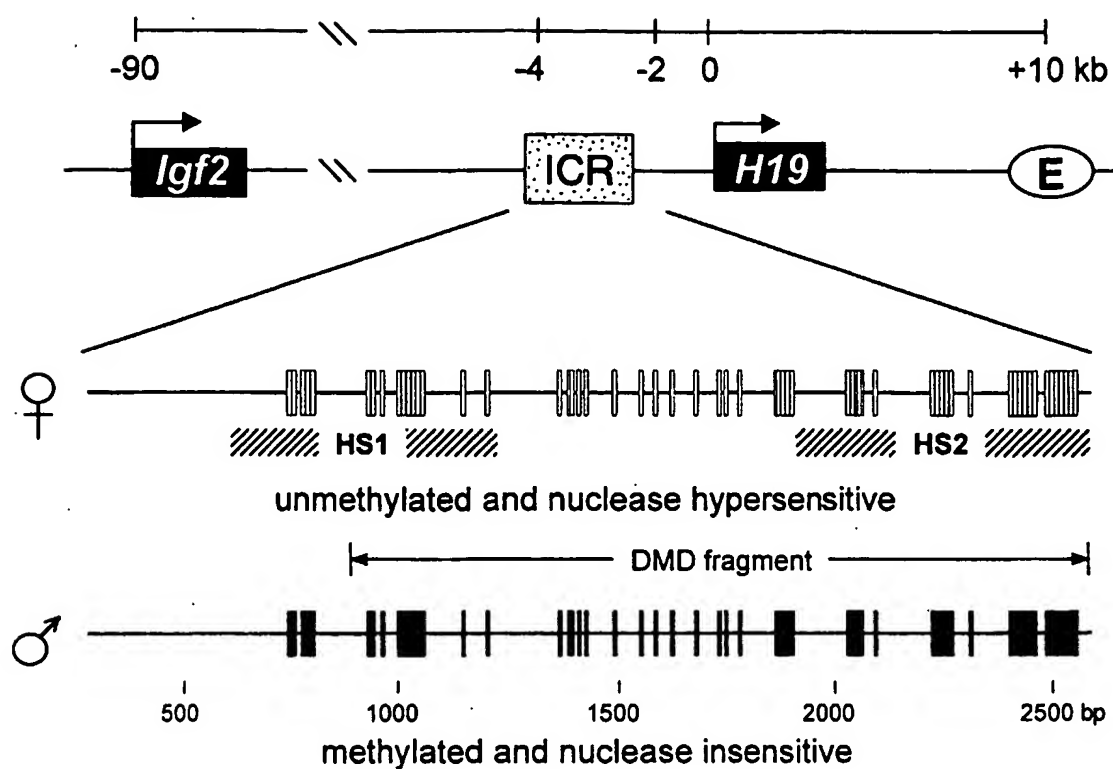


FIG. 8A

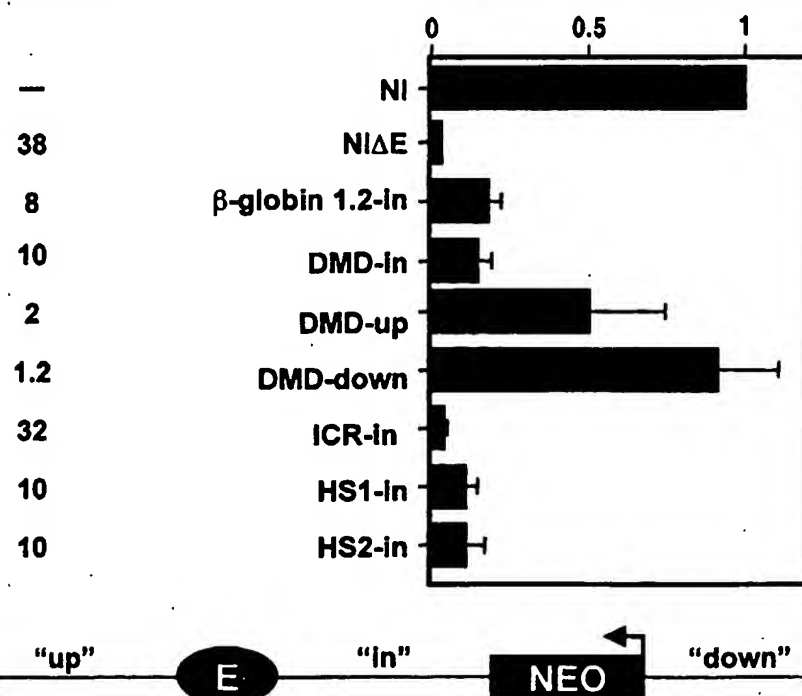
Fold ReductionRelative Neo^R Colonies

FIG. 8B

SUBSTITUTE SHEET (RULE 26)

Mouse
Ig α 2-H19

m1 GTTGTGGGTTTATACCGGGAGGTTGCCCGGTGGTGGCAGCAAAATCG SEQIDNO.84
m2 TTTGTGTG---TAAAGACACAGGTTGCCCGACCGCGCAGTGAATCT SEQIDNO.85
m3 TATGTGCAACAAGGAAAGGATGCTAACCGCGGTGGCAGCAATCTCC SEQIDNO.86
m4 GCTATA-GCTAAATGGACAGACGATGCCCGGTGGTGGCAGTAAATATAC SEQIDNO.87

Rat
Ig α 2-H19

r1 TTGTGTGGTTTAAACCGCGGAAAGTTGCCCGGTGGTGGCAGCAAAATC SEQIDNO.88
r2 TCCTTTGCGGTAAACACAGGCTGCCCGGTGGTGGCAGTGAATCG SEQIDNO.89
r3 TTGTGTGCAAGGAAATGATGTTAACCGCGGTGGCAGCAATCTCC SEQIDNO.90
r4 TGAATAATAGTAGATGGACAATAATGCCCGGTGGTGGCAGTAAATACCC SEQIDNO.91

Human
Ig α 2-H19

h1 GGTTGTAACGTGTGGAATCGAAAGTGGCGCGCGCGGCAGTGCAGGCT SEQIDNO.92
h2 GGTGTAGTTGTGGAATCGAAAGTGGCGCGCGCGGCAGTGCAGGCT SEQIDNO.93
h3 GGTGTAGTTGTGGAATCGAAAGTGGCGCGCGGCAGTGCAGGCT SEQIDNO.94
h4 GGTGTAGTTGTGGAATCGAAAGTGGCGCGCGGCAGTGCAGGCT SEQIDNO.95
h5 GGTGTAGTTGTGGAATCGAAAGTGGCGCGCGGCAGTGCAGGCT SEQIDNO.96
h6 GGTGTAGTTGTGGAATCGAAAGTGGCGCGCGGCAGTGCAGGCT SEQIDNO.97
h7 GGTGTAGTTGTGGAATCGAAAGTGGCGCGCGGCAGTGCAGGCT SEQIDNO.98

Ig α 2-H19
consensus

CCGCGNGGNGGCAG SEQIDNO.99

FIG. 9A β -globin F11 CCCAGGGATGTAATTACGTCCCTCCCECGGTAGGGGGGCAGCAGCGAGC SEQIDNO.100

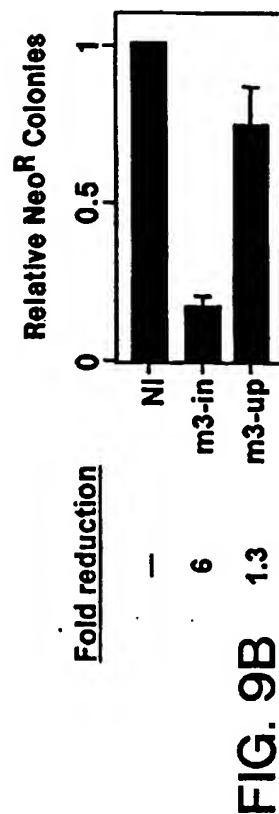


FIG. 9B

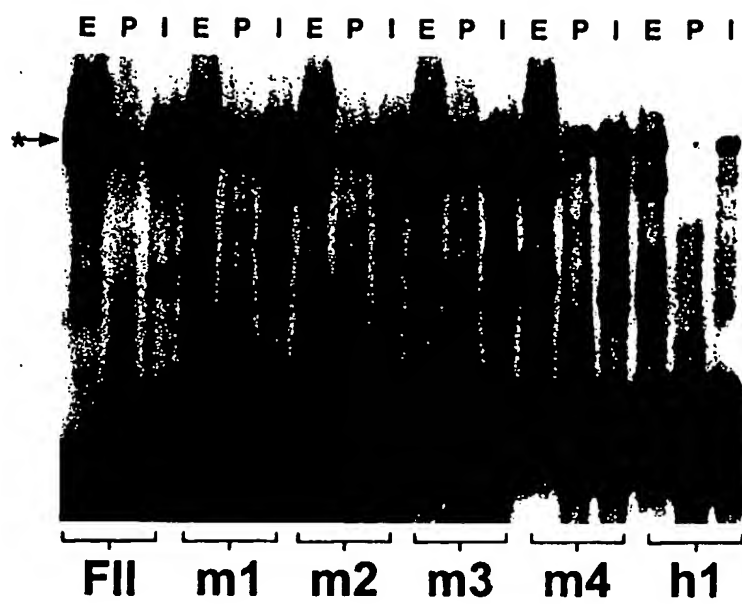


FIG. 9C

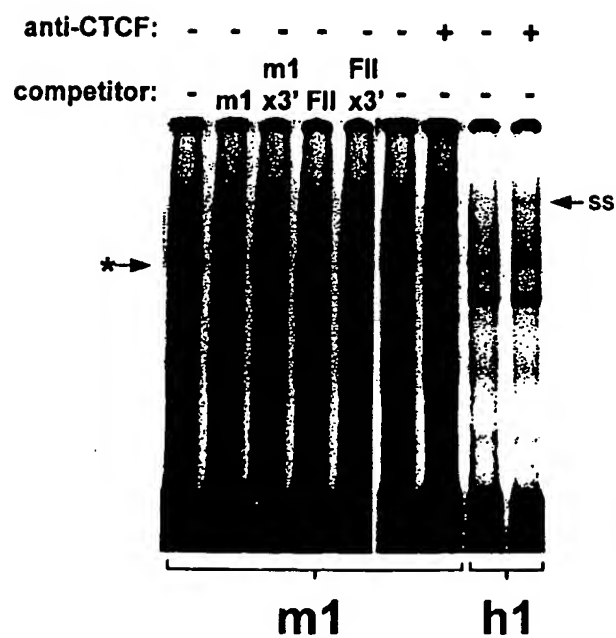


FIG. 9D

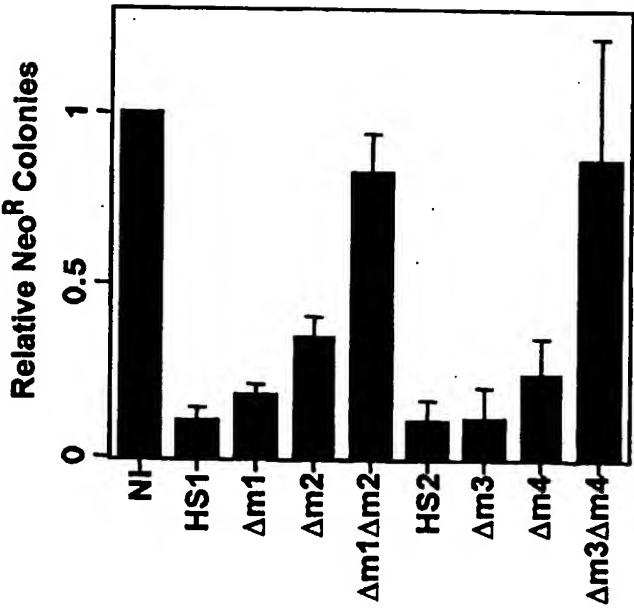


FIG. 10A

Fold Reduction

- 10 6 3 1.2 10 9 4 1.2

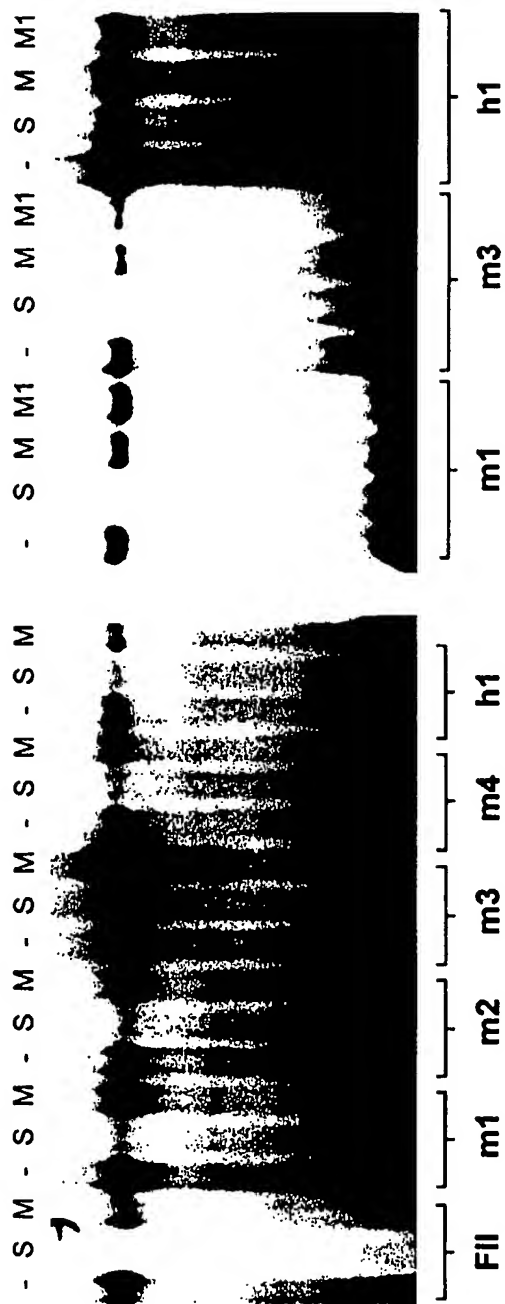
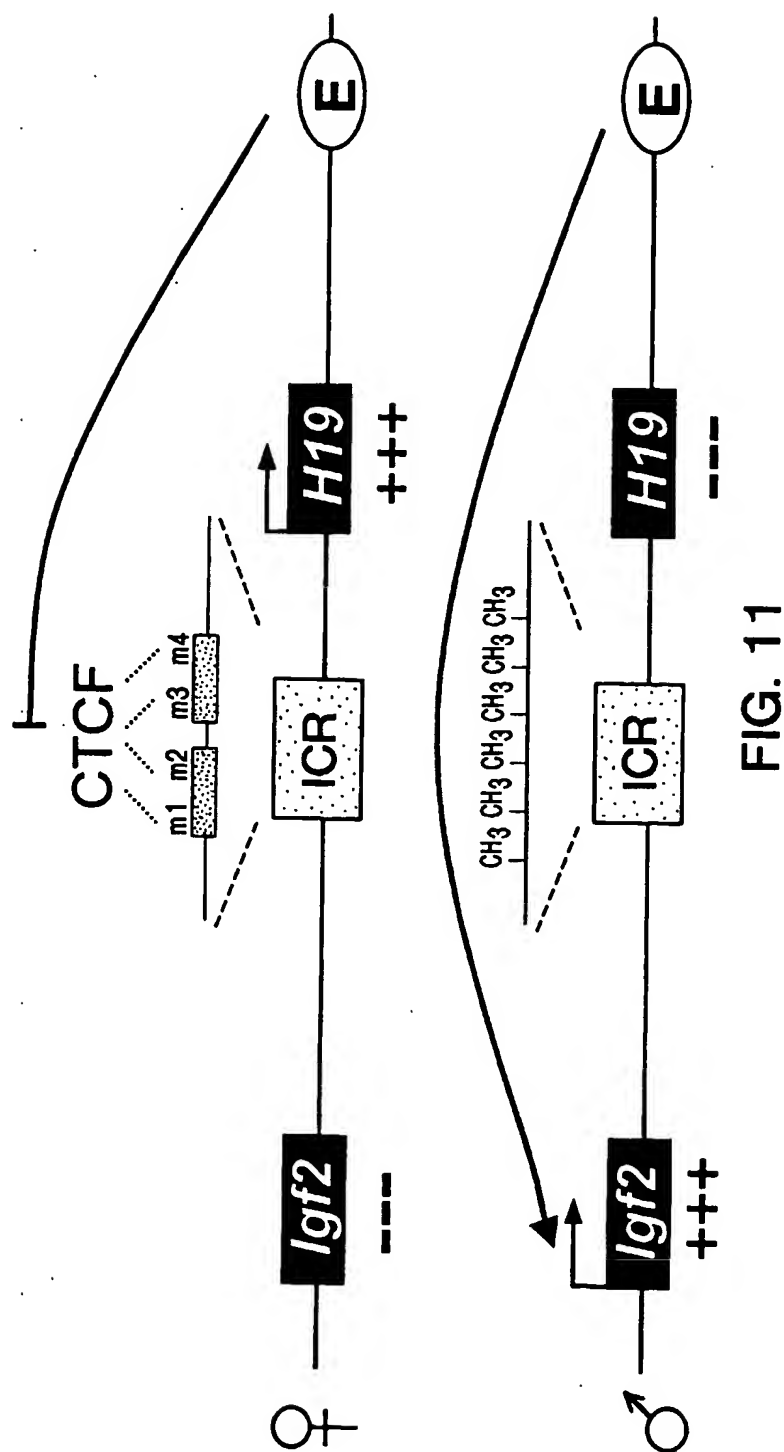


FIG. 10B



INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/10509

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/63 C12N15/85 C07K14/46 C12N9/10 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, STRAND, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHUNG, J.H. ET AL.: "Characterization of the chicken beta-globin insulator" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 94, no. 2, 21 January 1997 (1997-01-21), pages 575-580, XP002155512 cited in the application	1-11,13, 14
A	abstract page 575, column 1, line 1 - line 16 page 575, column 2, line 3 - line 14 page 576, column 2, line 27 -page 577, column 1, line 20 page 577; figure 3 page 577, column 2, line 1 - line 27 page 578; figure 4 page 578, column 2, line 29 -page 579, column 2, line 11 page 579; figure 5 -/--	12,15-27

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *g* document member of the same patent family

Date of the actual completion of the international search

15 December 2000

Date of mailing of the international search report

05/01/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Fuchs, U

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/10509

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	page 580, column 2, line 2 - line 6 ---	
X	FREVEL, M.A. ET AL.: "A potential imprint control element: identification of a conserved 42 bp sequence upstream of H19" TRENDS IN GENETICS, vol. 15, no. 6, June 1999 (1999-06), pages 216-218, XP004170206 cited in the application	17,18
A	the whole document	19-27
X	--- ZUBAIR, M. ET AL.: "Structure and Expression of the Mouse L23mrp Gene Downstream of the Imprinted H19 Gene: Biallelic Expression and Lack of Interaction with the H19 Enhancers" GENOMICS, vol. 45, no. 2, 15 October 1997 (1997-10-15), pages 290-296, XP002155513	16
A	abstract page 292 -page 293; figure 2 page 293, column 1, line 14 -page 294, column 1, line 22 page 294, column 2, line 25 - line 43 page 294, column 2, line 61 -page 295, column 2, line 15 page 295; figure 5 page 296, column 1, line 4 - line 12 ---	19-27
X	EMBL Database, Heidelberg, FRG Emrod accession number AF043428 06 January 1999 STADNICK, M.P. ET AL.: "Rattus norvegicus H19 gene, 5' flanking region" XP002155515	17
A	the whole document	19-27
P,X	--- BELL, A.C. ET AL.: "The Preotein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators" CELL, vol. 98, no. 3, 6 August 1999 (1999-08-06), pages 387-396, XP000926124 the whole document	1-12,15
A	--- WO 96 04390 A (GOVERNMENT OF USA, SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES) 15 February 1996 (1996-02-15) cited in the application the whole document --- -/--	1-15

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/10509

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 11972 A (THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK) 3 April 1997 (1997-04-03) the whole document ---	22-27
T	BELL, A.C. & FELSENFELD, G.: "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene" NATURE, vol. 405, no. 6785, 25 May 2000 (2000-05-25), pages 482-485, XP002155514 the whole document -----	1-27

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/10509

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9604390 A	15-02-1996	US 5610053 A AU 3150495 A	11-03-1997 04-03-1996
WO 9711972 A	03-04-1997	AU 7378196 A	17-04-1997

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☒ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.